University of Alberta

Issues in Process Control for the Syncrude Extraction Plant at their Mildred Lake Site

by

Eberhard Mueller

A thesis submitted to the faculty of graduate studies and research in partial fulfillment of the requirements for the degree of

Master of Science

in

Process Control

Department of Chemical Engineering

Edmonton, Alberta

Fall, 1996

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-18302-5

Canada

University of Alberta

Library Release Form

Name of Author: Eberhard Mueller

Title of Thesis: Issues in Process Control for the Syncrude Extraction Plant at their Mildred Lake Site

Degree: Master of Science

Year this Degree Granted: 1996

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, nether the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

169 Breckenwoods
51308 Range Road 224
Sherwood Park, Alberta
T8C 1H3 Canada

Dated: Sept. 30 1996

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Issues in Process Control for the Syncrude Extraction Plant at their Mildred Lake Site** submitted by **Eberhard Mueller** in partial fulfillment of the requirements for the degree of **Master of Science** in **Process Control**.

Professor S.L. Shah (Supervisor)
(Chemical and Materials Engineering)

Professor F. Forbes
(Chemical and Materials Engineering)

Professor M. Meng
(Electrical Engineering)

Dated: Sept. 26 1996.

To

my wife Pamela

and my children

Joe, Cindy, David, Chris, and Shane

Abstract

This thesis deals with several process control issues related to Syncrude's extraction process. Since startup in 1978, many changes have been made to the extraction control system and numerous new sensors were installed. The potential benefits of these changes and additions have not been fully realized as yet due to a lack of suitable real time models capable of integrating the information from the new sensors into the overall control strategy. The topics discussed here are part of an ongoing effort by Syncrude Research to improve the control strategy for the extraction plant to achieve better reliability and throughput for the process. These include the development of a method for on-line calibration and performance evaluation of composition analyzers applied to the plant feed composition analyzer, spectral analysis of process data and an example for the design and implementation of a suitable filter to improve control loop performance, process time delay estimation and empirical model identification.

Acknowledgments

I would like to express my sincere appreciation to my thesis supervisor, Dr. Sirish L. Shah, for his guidance during the preparation of this thesis. His suggestions, encouragement and patient teachings during the data analysis phase of the project proved to be invaluable.

I would also like to thank Syncrude Canada Ltd. and especially Syncrude Research's management, in particular Dr. John E. Clark, for their assistance and support in attaining this degree while being employed by them. I would be remiss not to say a special thanks to my Syncrude supervisor, Dr. Jan Czarnecki, for his tremendous support and his constant push during the final stages of this thesis, bringing it to a timely conclusion.

Finally, I would like to express my sincere thanks to all my many colleagues in the University of Alberta Chemical Engineering Process Control group for their help and numerous suggestions. In particular, I would like to thank Mr. S. Lakshminarayanan for his assistance during the PLS work for this thesis.

Above all, I would like to thank my wife, Pam, and my children for their assistance and patience over past years. Without their unwavering support and encouragement during my long hours of study, I would not have been able to complete this degree.

Contents

1. Introduction 1

1.1. Thesis outline 7

1.2. References 8

2. On-line sensor calibration and performance evaluation 9

2.1. Background 10

2.2. Principle of operation 11

2.3. Instrument description 12

2.4. Instrument calibration 13

2.5. Process description 15

2.6. Model based calibration 16

2.7. Data analysis and results 19

2.8. Effect of uncertainty in the process measurements 23

2.9. Feed rate evaluation 29

2.10. Conclusion 32

2.11. References 34

3. Spectral data analysis and filtering 35

3.1. Spectral analysis of feed rate signal 35

3.2. Downstream effects 38

3.3. Feed rate data filter design 42

3.4. Simulation results 46

3.5. Conclusion 47

4. Time delay estimation for primary separation vessel outputs 49

4.1. Use of cross-correlation for time delay estimate 52

4.2. Use of impulse response coefficient estimation for time delay determination 55

4.3. Transportation time based time delay estimation 61

4.4. Hybrid method for time delay estimation 64

4.5. Time delay variability 68

4.6. Conclusion 71

4.7. References 72

5. PLS analysis of extraction data 73

5.1. Objectives 73

5.2. Sensitivities of the PLS method 75

5.2.1. Effect of non steady state conditions on PLS results 75

5.2.2. Effect of time delay errors on PLS results 82

5.3. Time delays for the extraction data 84

5.4. Determination of data filters for the extraction data 86

5.5. PLS model for the extraction data 90

5.5.1. Four separate PLS models (MISO approach) 90

5.5.2. Single PLS model (MIMO approach) 97

5.6. PLS model validation 101

5.6.1. Validation for the MISO approach 101

5.6.2. Validation for the MIMO approach 105

5.7. Adaptive model 108

5.7.1. Simulation results using the adaptive model 109

5.8. Conclusion 112

5.9. References 115

6. Conclusions and future work 117

Appendix 120

References 132

List of Tables

2-1 Summary of calibration results 20

2-2 Summary of sensitivity analysis results 25

2-3 Comparison of different oil sand feed rate determination methods 31

3-1 Significant frequency points 43

3-2 Summary of simulation results for the period from August 19th at 0:00 to August 20th at 23:00, 1995 46

4-1 Summary of process inputs 51

4-2 Summary of process outputs 51

4-3 Time delay estimation results (minutes), based on cross-correlation 56

4-4 Time delay estimation results (minutes), based on impulse response 60

4-5 Summary of transportation delays up to the PSV feed distributor 62

4-6 Data and results for the froth layer time delay estimation 63

4-7 Matrix of time delays, in minutes, based on transportation delays only 64

4-8 Summary of PSV time delay estimates (minutes) based on feed slurry composition using cross-correlation 67

4-9 Summary of PSV time delay estimates (minutes) based on feed slurry composition using impulse response coefficients 68

4-10 Time delay variability due to changes in the froth layer thickness at a constant froth flow rate 70

5-1 PLS results for first simulation 77

5-2 PLS results for the second simulation using unfiltered data 80

5-3 PLS results for second simulation using filtered data 82

5-4 PLS results for time delay error analysis (unfiltered data) 83

5-5 PLS results for time delay error analysis (filtered data) 83

5-6 Total percent variance captured by PLS model for froth density as a function of time delay 85

5-7 Total percent variance captured by PLS model for froth flow rate as a function of time delay 85

5-8 Total percent variance captured by PLS model for froth dielectric as a function of time delay 85

5-9 Total percent variance captured by PLS model for middling density as a function of time delay 86

5-10 Time delays used for the PLS model 86

5-11 Summary of data filters for the PSV data based on autoregression of the residuals 87

5-12 Percent variance captured by the PLS model for froth density when all eleven latent variables are retained (MISO approach) 91

5-13 Percent variance captured by the PLS model for froth flow rate when all eleven latent variables are retained (MISO approach) 92

5-14 Percent variance captured by the PLS model for froth dielectric when all eleven latent variables are retained (MISO approach) 92

5-15 Percent variance captured by the PLS model for middling density when all eleven latent variables are retained (MISO approach) 93

5-16 Percent variance captured by the PLS model for froth density when five latent variables are retained (MISO approach) 94

5-17 Percent variance captured by the PLS model for froth flow rate when four latent variables are retained (MISO approach) 95

5-18 Percent variance captured by the PLS model for froth dielectric when six latent variables are retained (MISO approach) 95

5-19 Percent variance captured by the PLS model for middling density when four latent variables are retained (MISO approach) 95

5-20 Normalized regression coefficients for the PLS models (MISO approach) 96

5-21 Percent variance captured by the PLS model when all eleven latent variables are retained (MIMO approach) 98

5-22 Percent variance captured by the PLS model when six latent variables are retained (MIMO approach) 99

5-23 Normalized regression coefficients for the PLS models (MIMO approach) 100

5-24 Relative mean of residuals (%) when comparing the model output to actual measurements (MISO approach) 104

5-25 Relative mean of residuals (%) when comparing the model output to actual measurements (MIMO approach) 108

List of Figures

1-1 Schematic of the oil sand conditioning and primary recovery portion of a single train in Syncrude's extraction plant 5

2-1 Plant installation of Syncrude's Oil Sand Monitor 11

2-2 Optical measurement system for Oil Sand Monitor 13

2-3 Laboratory calibration setup for Oil Sand Monitor 14

2-4 Process schematic for the oil sand conditioning section of the extraction plant showing the location of the process measurements used in the on-line calibration 16

2-5 Comparison of the laboratory calibration curve and the two on-line calibration curves 21

2-6 Comparison of water content measured using the laboratory calibration curve and the two on-line calibration curves 22

2-7 Effect of a 10% relative bias in the feed rate on the on-line calibration curve 25

2-8 Effect of errors in the assumed reject density on the on-line calibration curve 26

2-9 Effect of a 2.5% relative bias in the PSV slurry density measurement on the on-line calibration curve 26

2-10 Effect of changes in the steam condensation factor inside the tumbler on the on-line calibration curve 27

2-11 Effect of a 5% bias in the flood water flow rate measurement on the on-line calibration curve 27

2-12 Effect of a 5% bias in the tumbler water flow rate measurement on the on-line calibration curve 28

2-13 Effect of a 0.5% absolute bias in the bitumen content determination on the on-line calibration curve 28

2-14 Comparison of plant feed conveyor and collection conveyor feed rates 29

2-15 Plot of the plant feed conveyor feed rate versus the collection conveyor feed rate 30

2-16 Comparison of the calculated feed rate to the two measured feed rates 32

3-1 Time series plot of oil sand feed rate 36

3-2 Frequency spectrum of oil sand feed rate 37

3-3 Tumbler water control loop schematic 39

3-4 Time series plot of tumbler water controller output 40

3-5 Time series plot of tumbler water flow 40

3-6 Frequency spectrum of tumbler water controller output 41

3-7 Frequency spectrum of tumbler water flow 41

3-8 Filter performance evaluation 43

3-9 Response of filter used for filtering the oil sand feed rate 44

3-10 Time series plot of filtered oil sand feed rate 45

3-11 Frequency spectrum of filtered oil sand feed rate 45

3-12 Comparison of tumbler water flow rates based on unfiltered and filtered feed rates 48

4-1 Block diagram of primary extraction circuit 49

4-2 Schematic for the primary extraction circuit showing the physical location of the sensors 50

4-3 Cross-correlation results for simulation of second order system with delay and coloured noise 53

4-4 Cross-correlation results for γ count and froth dielectric for September 4^{th}, 1995 54

4-5 Cross-correlation results for γ count and froth dielectric for September 5^{th}, 1995 54

4-6 Impulse response coefficients for a simulated second order system with delay and coloured noise 57

4-7 Impulse response coefficients for γ count and froth dielectric for September 4^{th}, 1995, using raw process data 57

4-8 Impulse response coefficients for γ count and froth dielectric for September 4^{th}, 1995, using normalized process data 58

4-9 Impulse response coefficients for γ count and froth dielectric for September 5th, 1995, using normalized process data 59

4-10 Block diagram of PSV input – output signals for PSV time delay estimation 65

4-11 Time delay variability due to changes in froth flow rate at a constant froth layer thickness 70

5-1 Step response for the three transfer functions used in the model 76

5-2 Plot of the three input signals used in the first simulation 77

5-3 Comparison of output signal for the first simulation and PLS model predictions 78

5-4 Plot of the three input signals used in the second simulation 79

5-5 Comparison of output signal for the second simulation and PLS model predictions using unfiltered data 79

5-6 Comparison of unfiltered and filtered residuals from the second simulation 81

5-7 Comparison of output signal for the second simulation and PLS model predictions using filtered data 81

5-8 Step response of the data filters for the froth density 88

5-9 Step response of the data filters for the froth flow rate 88

5-10 Step response of the data filters for the froth dielectric 89

5-11 Step response of the data filters for the middling density 89

5-12 Plot of regression coefficients (MISO approach) for eleven latent variable model for froth density, froth flow rate, froth dielectric and middling density 93

5-13 Plot of regression coefficients (MISO approach) for reduced number of latent variable model for froth density, froth flow rate, froth dielectric and middling density 96

5-14 Plot of regression coefficients (MIMO approach) for eleven latent variable model for froth density, froth flow rate, froth dielectric and middling density 99

5-15 Plot of regression coefficients (MIMO approach) for six latent variable model for froth density, froth flow rate, froth dielectric and middling density 100

5-16 Comparison of actual measurements to model predictions for September 5^{th} data set (MISO approach) 102

5-17 Comparison of actual measurements to model predictions for September 4^{th} data set (MISO approach) 102

5-18 Comparison of actual measurements to model predictions for September 6^{th} data set (MISO approach) 103

5-19 Comparison of actual measurements to model predictions for September 7^{th} data set (MISO approach) 103

5-20 Comparison of actual measurements to model predictions for September 9th data set (MISO approach) 104

5-21 Comparison of actual measurements to model predictions for September 5th data set (MIMO approach) 105

5-22 Comparison of actual measurements to model predictions for September 4th data set (MIMO approach) 106

5-23 Comparison of actual measurements to model predictions for September 6th data set (MIMO approach) 106

5-24 Comparison of actual measurements to model predictions for September 7th data set (MIMO approach) 107

5-25 Comparison of actual measurements to model predictions for September 9th data set (MIMO approach) 107

5-26 Comparison of actual measurements to model predictions for September 5th data set using the adaptive model 110

5-27 Comparison of actual measurements to model predictions for September 6th data set using the adaptive model 110

5-28 Comparison of actual measurements to model predictions for September 7th data set using the adaptive model 111

5-29 Trajectory of model update parameter for September 5th data set 112

5-30 Trajectory of model update parameter for September 5th data set 113

5-31 Trajectory of model update parameter for September 5th data set 113

List of Abbreviations

DCS	Digital Control System
EAPS	Extraction Auxiliary Production System
FFT	Fast Fourier Transform
IIR	Infinite Impulse Response
MIMO	Multiple Input / Multiple Output
MISO	Multiple Input / Single Output
MLR	Multiple Linear Regression
NIRA	Near Infrared Reflectance Analysis
PLS	Partial Least Squares
PSV	Primary Separation Vessel

List of Symbols

B	Bitumen fraction in oil sand
CE	Control effort
$Ch1$	Water measurement signal
$Ch2$	Water reference signal
FR	Oil sand feed rate
FW	Flood water volume flow rate
h_{Froth}	Height of froth layer
m_B	Mass flow rate of bitumen
m_B^{Slurry}	Mass flow rate of bitumen in slurry
m_C	Mass flow rate of caustic to tumbler
m_C^{Slurry}	Mass flow rate of caustic in slurry
$\overline{m}_{Froth}$	Average froth mass flow rate
m_R	Mass flow rate of reject
m_S	Mass flow rate of solids
m_S^{Slurry}	Mass flow rate of solids in slurry
m_W	Mass flow rate of water
m_W^{Slurry}	Mass flow rate of water in slurry
P	Period of a signal
$P_{Sampling}$	Sampling period
q	Backward shift operator

r_{PSV}	Radius of PSV
SR	Steam mass flow rate
TW	Tumbler water volume flow rate
u	Process input
W	Water fraction in oil sand
x	Input data value
$\bar{x}$	Average input value
x^*	Scaled input data value
x_B^{Slurry}	Mass fraction of bitumen in slurry
x_C^{Slurry}	Mass fraction of caustic in slurry
x_S^{Slurry}	Mass fraction of solids in slurry
x_W^{Slurry}	Mass fraction of water in slurry
X_1	Intercept for water calibration curve
X_2	Slope for water calibration curve
y	Output signal
$\hat{y}$	Predicted output
α	Steam condensation factor
β	Regression coefficient
λ	PLS model update parameter
ρ_B	Density of bitumen
$\bar{\rho}_{Froth}$	Average froth density
ρ_R	Density of reject

ρ_S	Density of solids
ρ_{Slurry}	Density of PSV feed slurry
ρ_W^F	Density of flood water
ρ_W^{Slurry}	Density of water in PSV feed slurry
ρ_W^T	Density of tumbler water
σ_x	Variance of input data set

1. Introduction

Syncrude's predecessor, then part of Royalite Oil Company, started research into bitumen extraction from the Athabasca oil sands in the mid 1950's. The company was incorporated under the Syncrude name as a separate entity in 1964 and continued it's research efforts until it received permission to build a production facility in 1969. After several years of engineering effort, construction commenced in late 1973 and the plant went into operation in 1978.[1]

Currently, Syncrude is the world's largest producer of synthetic crude oil, the largest single source of oil in Canada and Canada's second largest oil producer with an average production of over 200,000 barrels (1 barrel equals 159L) per day of high quality synthetic crude oil. This represents approximately 20% of light and medium crude oil produced in Canada. In 1995 Syncrude produced 73.9 million barrels of crude, more than 12% of Canada's total petroleum needs, at a cost of $13.69 per barrel, a 37% reduction since 1981. The expected production for 1996 is 75 million barrels.

The Company provides an estimated 15,000 direct and indirect jobs across Canada and is the country's largest individual employer of Aboriginal people. Syncrude employs about 3,600 people directly in addition to some 1,000 contractors. It's corporate headquarters are located in Fort McMurray (440km northeast of Edmonton, Alberta) and it's production facility is 40km north of Fort McMurray.

The Syncrude production facility consists of three major strategy areas: mining, extraction and upgrading. Oil sand is mined in a large open pit mine (current size: 5km by 7km by 60m deep) at a present rate of over 141 million tonnes per year using a combination of draglines and bucketwheels. Oil sand is transported over a

series of conveyor belts totaling 50km in length to the processing facility. The Extraction plant processes in the neighborhood of 360,000 tonnes of oil sand per day, separating the bitumen from the sand. The upgrading facility converts the black, tar-like bitumen to low sulphur synthetic crude oil, similar in appearance and viscosity to standard engine oil. In addition to the production units, the company operates a large utilities plant providing steam (1,700,000 $\frac{kg}{hr}$ at 6,400kPa and 500°C), generating electricity (268 megawatt capacity) as well as producing air for plant instrumentation and nitrogen for some of the process units. A very important part of the operation is the environmental department. With an annual budget of over $6 million, it has been in existence since the mid 1960's, almost ten years before construction was started. It's mandate is to carry out research into minimizing the environmental impact of the operation and to develop and implement efficient methods of land reclamation. Since the reclamation phase of their work started in 1979, over 1.3 million trees seedlings were raised in Syncrude's greenhouses and planted. In 1993, 48 wood bison, a species that used to inhabit this area, but had been absent for many decades, were successfully introduced into the newly reclaimed land from Syncrude's mining operation.

The company engages in extensive, high-tech research and development activities with an annual expenditure of over $24 million.[2] The Edmonton based research facility is the third largest private-sector research program in Western Canada and ranks in the top 25 R&D spenders in this country. The Research department employs just under 90 full-time staff and has extensive links with the University of Alberta and other Canadian universities as well as to private and government research facilities. Research carried out at the Edmonton facility encompasses eight diverse fields. These include studies in the following areas:

- separation processes

- measurements and analysis
- bitumen chemistry
- colloid and surface science
- analytical chemistry
- materials research
- tailings management
- slurry transport

This thesis deals with a few process control related aspects of Syncrude's extraction facility: basic control loop analysis, on-line instrument calibration and process model development.

Since production started in 1978, the extraction plant was operated according to a set of operating guidelines which were refined over the years as operating experience increased. The impact that the changes to these guidelines have had on the operation was difficult to assess, since no on-line measurements of stream composition existed. For example, the goal of the primary extraction process is to extract the maximum amount of bitumen from the oil sand feed and deliver the best possible quality of bitumen froth to the downstream units. Yet no sensors were available to measure any of the properties of the oil sand feed to the plant, the water and solids content of the froth, the amount of froth produced, or the amount of bitumen lost through the waste streams. Over the years, on-line sensors for oil sand composition, froth composition and froth flow rate, all developed at the research department, have been installed.

In 1990 the original control room instrumentation was replaced by a modern, computer based control system. The commissioning of this system made it possible to consider introducing an automatic control strategy for the extraction plant. It was anticipated that computer based control would make it possible to reduce the number

of upset conditions, make more efficient use of costly resources, obtain more uniform product quality and enable higher throughput. The first step in developing such a strategy is to attempt a mathematical model of the major process units. This model must be able to make maximum use of the few on-line process data points that are now available. To date, a model that could be used for real time operator assistance does not exists.

The oil sand processed by Syncrude is a mixture of bitumen, water, and solids of varying proportions. Bitumen content can vary from just a few percent by weight to as high as 15% or 16%, with an average of somewhere between 9% and 11%. Oil sand with bitumen content of less than 6% is discarded in the mine and does not reach the extraction plant under normal circumstances. Water content ranges from 3% to 8% by weight. The remainder of the oil sand, 80% to 90%, is solids. The type of solids contained in the oil sand significantly impacts on the performance of the extraction circuit. Quartz and silica particles with diameters larger than 40 μm have no effect on the process. Clay particles on the other hand can have a serious impact on the amount of bitumen recovered and on the quality of the froth produced, especially if their concentration in the middling layer in the Primary Separation Vessel (PSV) gets too high.

The Syncrude extraction plant consists of four identical, parallel processing streams ('trains'). Each train can be divided into four sections: oil sand conditioning, primary recovery, secondary recovery and tailings disposal. The work presented in this thesis focuses on a few aspects of the first two sections of the extraction process. Figure 1-1 represents a schematic view of one of the trains up to and including the PSV. Oil sand is delivered from the mine via conveyor belts at a nominal rate of $1250 \frac{kg}{sec}$. It is loaded into a 30m long, cement kiln type tumbler where it is mixed

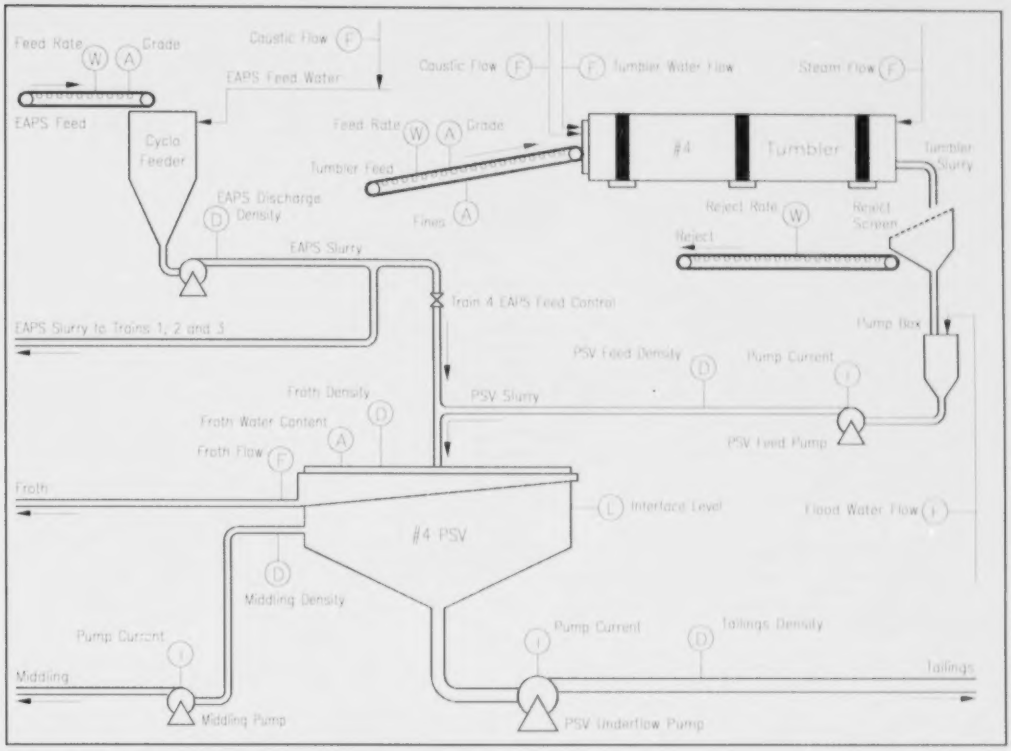

Figure 1-1: Schematic of the oil sand conditioning and primary recovery portion of a single train in Syncrude's extraction plant.

with caustic (NaOH) and hot water and heated to about 80°C with steam. The oil sands slurry exiting the tumbler is screened to remove large rocks and undigested oil sand lumps, diluted with additional hot water and pumped into the PSV.

The PSV has a diameter of 18.9m and a volume of over 1,700,000L. It has three distinct zones. At the top is a froth layer with a nominal thickness of 0.8m. The conical shaped bottom of the vessel contains mainly coarse sand that is removed through the tailings line. The center portion is made up of the middling layer. The diluted slurry is introduced into this layer and it is the region where the separation between the bitumen and the sand occurs.

Once the diluted slurry is introduced into the center of the PSV, the heavy

sand settles almost immediately to the bottom of the cone shaped vessel and is removed as primary tailings. The bitumen contained in the slurry rises to the top of the vessel and overflows into a collection trough. In order to balance the flow into the PSV with the outflow a third stream, the middling stream, is withdrawn from the central periphery. It contains mostly water, fine solids and some bitumen.

The sensors indicated in the Figure 1-1 represent all the relevant measurements currently available for this portion of the process. On the feed conveyor a weigh scale ('Feed Rate') records the instantaneous oil sand feed rate; the Oil Sand Analyzer ('Grade') measures the oil sand bitumen and water content; and the 'Gamma' analyzer ('Fines') gives some information about the clay content in the feed. It measures the naturally emitted gamma radiation from one of the potassium isotopes (K^{40}) which are associated with the oil sand clays. Three flow meters report the volumes of caustic, tumbler water and flood water added to the process. A density gauge on the discharge end of the feed pump gives diluted slurry density. A second weigh scale ('Reject Rate') on the reject conveyor records the amount of material that could not be processed. Probes located in the froth layer at the top of the separation vessel give an estimate of froth water content and froth density. In addition, on-line density is available for the middling and tailings streams. The position of the interface between the middling layer and the froth is also recorded ('Interface Level'). Finally, a flow meter is located on the froth discharge line from the PSV. The part of the extraction plant described so far represents the original portion of the process (base plant).

In the early 1990's, the Extraction Auxiliary Production System (EAPS) was built and brought on-line to boost processing capacity. It differs from the base plant in the oil sand conditioning step. The oil sand feed is passed through a crusher to reduce the size of the oil sand lumps and any rocks that are present. The crushed oil

sand is then mixed with hot water and caustic in a cyclofeeder and the resulting slurry is pipelined over several kilometers and pumped directly into the PSV. Only minimal measurements are available for this portion of the plant, which poses considerable difficulties in the development of process models.

The Extraction Plant is controlled by a Honeywell TDC3000 control system. Data collected by the DCS is transferred at one minute intervals to a plant wide database on a dedicated VAX computer. Data from this data base was for the investigations presented in this thesis.

1.1. Thesis outline

The purpose of this thesis was to apply different data analysis tools to illustrate how useful information can be extracted from routine plant operating data. The work presented here is a small part of a much larger ongoing effort to develop real-time optimization and operator assistance tools that aim at improving both the bitumen recovery efficiency and product quality in the extraction plant.

Chapter 2 explores the feasibility of using the current plant database information to calibrate on-line sensors and evaluate their field performance. Two specific examples are given: calibration of the Oil Sand Monitor, which measures oil sand composition as it enters the extraction plant, and the performance evaluation of the oil sand feed weighing system. Also included is an analysis of the accuracy which can be expected from such a method based on accuracy estimates for the individual measurements used.

In Chapter 3, spectral data analysis is applied to identify noise components within the process measurements and their downstream effects. An example is given, where the design of a proper data filter for a measurement, based on the results of the

spectral analysis, has dramatically improved the performance of a control loop.

An important step in the development of a process model, as well as in controller design, is the estimation of process time delays (dead times). Chapter 4 describes four methods used to estimate the time delays for the extraction process up to and including the PSV.

An empirical model for the PSV is developed in Chapter 5 using Partial Least Squares (PLS) analysis. The applicability of this method to this particular data set as well as its limitations are examined. Also included is a simulation technique, based on the PLS model, that attempts to compensate for slow, long term changes in plant behavior as well as for unmeasured process inputs due to periodic changes in the plant flow configuration which are not recorded in the plant database.

1.2. References

[1] Micheal Lupien, Ed., "Everything you ever wanted to know about Syncrude" Syncrude Canada Ltd., Public Affairs Department, Fort McMurray, Alberta.

[2] "Celebrating the science behind the sands" Syncrude Canada Ltd., Syncrude Research Centre, Edmonton, Alberta.

2. On-line Sensor Calibration and Performance Evaluation

Without sensors, any form of process control is impossible. For these sensors to be useful, their output signal (usually some electrical quantity) must usually be transformed into a physically meaningful quantity (generally engineering units). This transformation is accomplished using experimentally derived calibration curves. The accuracy of the sensors depends heavily on how closely the experimental conditions in the laboratory can be matched to the actual service conditions. For simple sensors, like a temperature sensor for example, matching of calibration and service conditions is not very critical and when this sensor is installed in the field, there is a high degree of confidence that the reported temperature will be indicative of the actual process temperature. Another class of sensors, such as flow sensors, require more closely matched laboratory and service conditions. If the process fluid to be measured is a single phase material with well defined and relatively constant physical properties, matching of laboratory condition to the process should not pose any significant problems, and the resulting calibration will yield representative results. For many advanced control applications however, measurements of simple stream properties are not sufficient. These applications often rely on sophisticated on-line composition analyzers. Performing calibrations on this class of sensors can be quite difficult. In most cases, it is not possible to match the laboratory conditions to the process, which raises the question of how accurately the readings obtained from the unit reflect stream composition when it is on-line. In addition, these sensors often exhibit calibration drift over time and require periodic re-calibration. It would be advantageous to have some method of verifying the performance of such sensors, and possibly re-calibrating them, while they are on-line.

Syncrude's extraction process employs a composition sensor on the oil sand

feed stream to the plant, commonly referred to as the Oil Sand Monitor. Even though this sensor has been operating in the plant for quite some time, it has not been possible to verify it's on-line performance until recently. In the remainder of this chapter, the on-line performance of the Oil Sand Monitor will be examined.

2.1. Background

In 1983 the first version of the Oil Sand Monitor was installed in the extraction plant on the feed conveyors. This instrument was designed to measure the bitumen content of the oil sand feed on-line. The unit was a commercially available instrument intended to measure the presence or absence of an oil film on top of effluent water. It measured the intensity of two specific wavelengths of light reflected from the water surface. The wavelengths were selected by two light filters mounted in a rotating filter wheel. The instrument responded to a change in the ratio of the intensities of these two wavelengths. Substantial in-house modifications were made to the instrument in order to measure bitumen content in oil sand. The information from the Oil Sand Monitor became a vital part of the control strategy for extraction. The instrument was replaced by a newer version in 1990. Even though this version of the instrument was optically and mechanically designed to use up to eight light filters, no provisions were made in the electronics to process information from more than two. Laboratory tests have shown that, if more wavelengths were available, the instrument would be capable of measuring other oil sand components in addition to bitumen. A project was carried out aimed at replacing the analog electronics supplied by the instrument manufacturer with microprocessor based digital electronics capable of measuring up to eight signal channels. Several benefits were realized as a result of this project. The signal to noise ratio was improved by several orders of magnitude, the instrument drift present in the earlier versions was greatly reduced and the water

content of oil sand could now be measured. A prototype instrument was installed in the extraction plant in 1994 and evaluated over an extended period of time. In early 1996 all instruments in the plant were replaced with the new version.

2.2. Principle of operation

The oil sand analyzers are based on the principals of Near Infrared Reflectance Analysis (NIRA). The oil sand is illuminated by a 250W halogen light source as it moves past the instrument on the feed conveyor (Figure 2-1). Some of the light from this source is diffusely reflected back towards the analyzer, where the intensity of the reflected light at several wavelengths is measured. One wavelength was selected to be in a region of the spectrum where the bitumen does absorb (measurement signal), the other is in a region where the bitumen does not absorb (reference signal). The

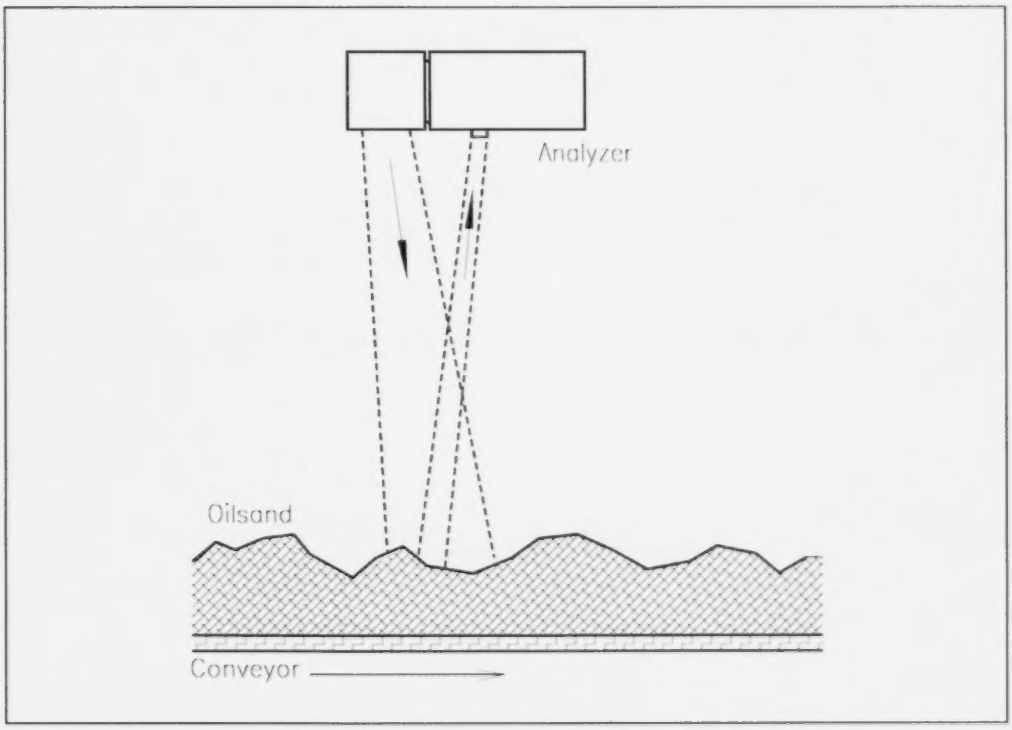

Figure 2-1: Plant installation of Syncrude's Oil Sand Monitor.

ratio of the intensities of the reflected light at these two wavelengths indicates the amount of bitumen in the oil sand. Similarly, two different wavelengths were chosen (one as the measurement signal and one as the reference signal) in such a way that their ratio gives a measure of the water content in the oil sand.

The instrument is designed to operate at a distance of approximately 55 inches from the oil sand surface and monitors a circular area with a 2 inch diameter. The accuracy is estimated at plus or minus 0.5% bitumen by weight (absolute).

2.3. Instrument Description

The instrument is divided into two sections. A transmitter housing and a receiver housing. The two sections are physically separated to prevent overheating of the receiver by the lamp. The transmitter consists of a 250W, 115VAC halogen lamp mounted at the focal point of a parabolic reflector.

The receiver houses the mechanical, optical, and electronic components of the instrument. The mechanical and optical components are closely related and will be described together. The light enters the receiver through an infrared transmitting lens, which focuses the light beam onto the infrared photo detector. A rotating filter wheel, with provisions for up to eight optical interference filters, is located between the receiver lens and the detector (Figure 2-2). The purpose of this filter wheel is to 'chop' the incoming light beam and pass light at specific wavelengths to the sensing element. There are seven optical filters and one blank installed in the filter wheel. The objective of the blank is to block all light to the detector so that an electronic background reference can be determined. The filter wheel rotates at approximately 3300 revolutions per minute.

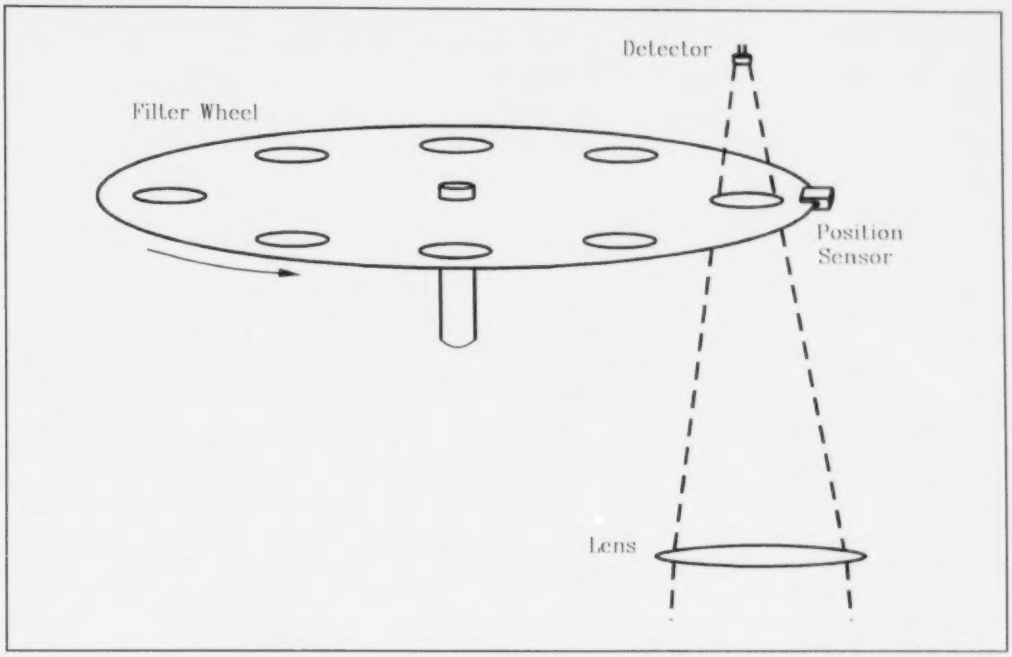

Figure 2-2: Optical measurement system for Oil Sand Monitor.

The electronics for the instrument consist of an analog section, for pre-processing the detector signal, an analog-to-digital converter and a digital section for data processing and filter wheel synchronization. Data processing consists of separating the eight serially acquired signals from each revolution of the filter wheel into their respective channels. The electronic background reference signal is then subtracted from each of the seven measurement channels and, since NIRA methods inherently result in quite noisy measurements, each channel is averaged over approximately 25 seconds.

2.4. Instrument calibration

The instruments are initially calibrated in the laboratory using actual oil sand samples of known composition. Secondary standards were developed for use in the field to adjust for instrument drift during normal operation. Both calibration methods

use static samples, whereas the process measurement is on a sample moving at 620 feet per minute. During laboratory calibration, a pan of oil sand is placed under the instrument (Figure 2-3). The surface of the oil sand in the pan is leveled and does not vary significantly from one sample to the next. The composition of each sample has been determined previously by an analytical reference method. In each case, the instrument averages about 1375 light intensity readings from the same area of the oil sand sample prior to reporting a composition value. Any difference observed from one sample to the next is solely due to changes in oil sand composition. When the instrument is taking measurements on-line, the measurement surface changes continuously. Neither the distance to the instrument, nor the surface orientation and topography can be controlled. Since the intensity measurements for the different wavelengths of light are taken sequentially, not even the measurement and reference signals used in a single ratio calculation are from exactly the same area. Sampling the

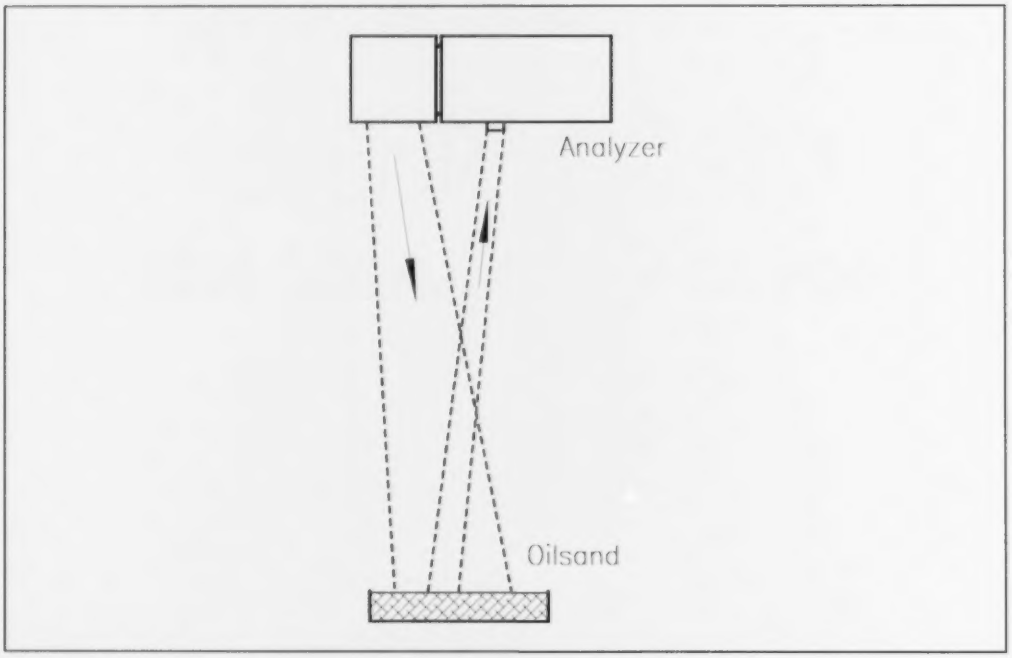

Figure 2-3: Laboratory calibration setup for Oil Sand Monitor.

oil sand on the conveyor belt during normal operation and comparing the laboratory analysis to the instrument reading is impossible due to the speed of the moving oil sand. One 25 second average reading of the instrument represents, at nominal feed rates, approximately 35,000kg of oil sand.

Laboratory calibration of the instrument for bitumen content in the oil sand is relatively easy, since it changes very slowly when the sample is handled and exposed to the air. Calibrating for water content is a different matter. The oil sand will lose water every time it is transferred from the sample container to the pan and also while it is under the instrument due to the heating effect from the intense illumination by the instrument's light source. This introduces an uncertainty into the calibration curve for this measurement.

2.5. Process description

The following is a brief description of the process as it pertains to the model developed in the next section. A process schematic showing all measurements relevant to the model is presented in Figure 2-4. Oil sand is fed to the tumbler via two conveyor belts in series. The first one is the collection conveyor and the second one is the plant feed conveyor. Each of these two belts have independent measurements of feed rate. Hot water (tumbler water) and an aqueous caustic solution are added to the oil sand as it enters the tumbler. The amount of caustic solution added is very small compared to all the other flow rates (<0.01% by weight of total flow) and was omitted from the model. Steam is injected at the discharge end of the tumbler, some of which condenses and adds to the water in the tumbler and the rest of the steam vents into the atmosphere. The main purpose of the steam addition is to raise the temperature of the oil sand and water mixture. The heavy slurry

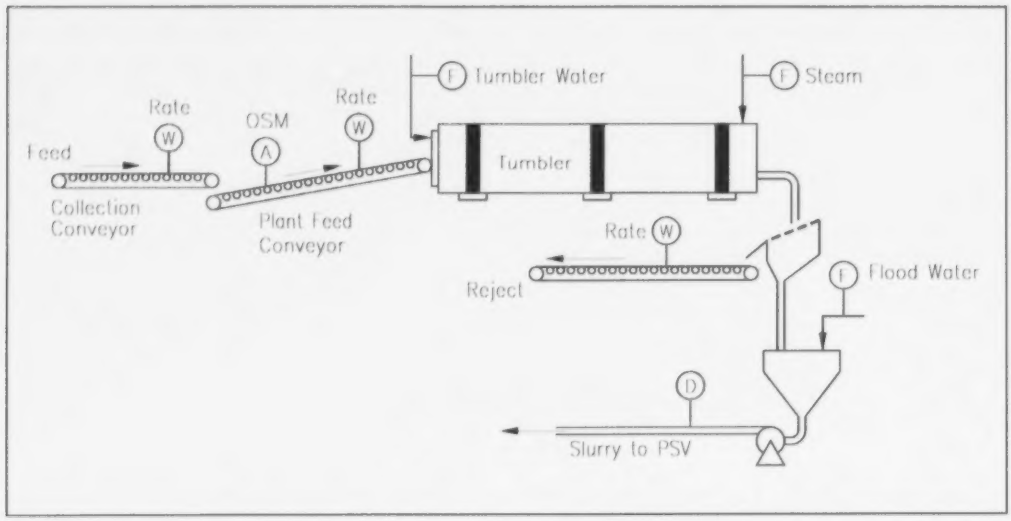

Figure 2-4: Process schematic for the oil sand conditioning section of the extraction plant showing the location of the process measurements used in the on-line calibration.

discharged from the tumbler is passed over a set of vibrating screens, which remove any large, undigested oil sand lumps and rocks. The material that does not pass through the screens is removed as reject. The rest of the stream is further diluted with hot water (flood water) in the pump box to form the feed for the primary separation vessel (PSV), referred to as PSV slurry.

2.6. Model based calibration

Oil sand contains three main constituents: bitumen, water and solids. As long as only bitumen could be measured it was impossible to evaluate the sensor performance, such as instrument drift for example, using any type of material balance model. With the addition of the water measurement two out of the three components are now known making it possible to perform an on-line performance evaluation for this instrument.

Due to the difficulties in the laboratory calibration for the water measurement, an attempt was made to derive an on-line, model based water calibration curve. The hypothesis being, that if this calibration curve is close to the one obtained in the laboratory, the instrument is giving valid results on-line and the lab calibration procedures are valid.

The model encompasses the feed system, the tumbler, the reject system and the pump box. The definition of the notation used in the model equations is as follows:

m_W = mass flow rate of water
m_B = mass flow rate of bitumen
m_S = mass flow rate of solids
m_R = mass flow rate of reject
ρ_{Slurry} = PSV slurry density
ρ_W^{Slurry} = density of water in PSV slurry
ρ_W^T = density of tumbler water
ρ_W^F = density of flood water
ρ_B = density of bitumen in PSV slurry
ρ_S = density of solids in PSV slurry
ρ_R = density of reject
W = water fraction in oil sand
B = bitumen fraction in oil sand
FR = oil sand feed rate
TW = tumbler water volume flow rate
FW = flood water volume flow rate
SR = steam mass flow rate
α = steam condensation factor
$Ch1$ = water measurement signal
$Ch2$ = water reference signal
X_1 = intercept for water calibration curve
X_2 = slope for water calibration curve

The material balance model used here is based on the PSV feed slurry density, which can be expressed as a function of process input and output rates:

$$\rho_{Slurry} = \frac{m_W + m_B + m_S - m_R}{\frac{m_W}{\rho_W^{Slurry}} + \frac{m_B}{\rho_B} + \frac{m_S}{\rho_S} - \frac{m_R}{\rho_R}} \tag{2.1}$$

Where the mass flow rates of the individual constituents are as follows:

$$\begin{aligned} m_W &= (W)(FR) + \rho_W^T(TW) + \alpha(SR) + \rho_W^F(FW) \\ m_B &= (B)(FR) \\ m_S &= (1 - W - B)(FR) \end{aligned} \tag{2.2}$$

Substitution of these mass flow rates into the Equation 2.1 yields:

$$\rho_{Slurry} = \frac{(W)(FR)+\rho_W^T(TW)+\alpha(SR)+\rho_W^F(FW)+(B)(FR)+(1-B-W)(FR)-m_R}{\frac{(W)(FR)+\rho_W^T(TW)+\alpha(SR)+\rho_W^F(FW)}{\rho_W^{Slurry}}+\frac{(B)(FR)}{\rho_B}+\frac{(1-B-W)(FR)}{\rho_S}-\frac{m_R}{\rho_R}} \quad (2.3)$$

Based on information gathered during laboratory calibration, the water fraction can be expressed as:

$$W = X1 - X2\frac{Ch1}{Ch2} \quad (2.4)$$

Replacing the water fraction, W, in Equation 2.3 with Equation 2.4 and rearranging the resulting expression into the form $\beta_1 X_1 - \beta_2 X_2 = \gamma$, allows it to be solved for $X1$ and $X2$ using linear regression. This leads to the following expression:

$$\begin{aligned}&(FR)\left(\frac{1}{\rho_W^{Slurry}}-\frac{1}{\rho_S}\right)X1-\frac{(FR)(Ch1)}{(Ch2)}\left(\frac{1}{\rho_W^{Slurry}}-\frac{1}{\rho_S}\right)X2\\&=\frac{\rho_W^F(FW)+\rho_W^T(TW)+\alpha(SR)+(FR)-m_R}{\rho_{Slurry}}\\&-\frac{\rho_W^T(TW)+\alpha(SR)+\rho_W^F(FW)}{\rho_W^{Slurry}}-\frac{m_R}{\rho_R}+\frac{(B)(FR)}{\rho_B}+\frac{(1-B)(FR)}{\rho_S}\end{aligned} \quad (2.5)$$

Several assumptions had to be made to enable the development of this model:

- Since no reject composition data is available on-line, the long term average based on lab analyses was used: $\rho_R = 2.0\frac{kg}{L}$.
- Bitumen density is constant at: $\rho_B = 1.008\frac{kg}{L}$.
- Solids density is constant at: $\rho_S = 2.65\frac{kg}{L}$.
- The steam condensation rate inside the tumbler was fixed at 75% resulting in: $\alpha = 0.75$.

- The tumbler water temperature was assumed to be constant at 90°C: $\rho_W^{TW} = 0.962 \frac{kg}{L}$

- The flood water temperature was taken as 80°C: $\rho_W^F = 0.972 \frac{kg}{L}$.

- The temperature used for the PSV slurry was also 80°C: $\rho_W^{Slurry} = 0.972 \frac{kg}{L}$.

2.7. Data analysis and results

Selection of a good data set for the calibration is critical regardless of whether it is for an on-line calibration or a laboratory calibration. When samples are chosen for a laboratory based calibration, the widest possible composition range is selected in order to adequately cover the entire measurement range of the instrument. The criteria for plant data selection on the other hand is that the plant was operating in a stable manner and under control over the entire selected time period. Unfortunately, this is only the case when the feed composition is relatively stable. Consequently, the instrument will only be calibrated over a very small portion of its operating range. In this case, small disturbances in the data set will result in large biases for the calibration constants.

The sample set selected for the lab calibration consisted of 65 oil sand samples ranging in water content from 0.16% to 9.71% as determined by Syncrude's standard analytical method for oil sand assay. Each sample was placed under the Oil Sand Monitor and the readings for channel 1 (water measurement wavelength) and channel 2 (water reference wavelength) were recorded. Care was taken during the handling and measurement steps to minimize the evaporation of water. The calibration curve was derived by linear regression of the water content and the ratio on the two channel readings for each sample.

The data set for the on-line calibration was taken from the plant data base. The selection criteria was three consecutive days of operation without feed interruptions and upsets. The criteria could not quite be met and a sample set of 3536 samples covering 59 hours of operation from July 29th, 1995 at 13:00 to August 1st, 1995 at 0:00. At each one minute sample step, all ten model parameters were recorded. One of these, the oil sand feed rate, had redundant measurements which unfortunately differed significantly from each other. Since the feed rate is a key variable in the model, the calibration calculations were carried out twice, once using the plant feed conveyor values and once using the collection conveyor values. The result are summarized in Table 2-1.

Upon initial inspection, all three results look significantly different. However, when the calibration curves are plotted (Figure 2-5), it becomes apparent that the calibration based on the collection conveyor feed rate is much closer to the lab calibration curve than is the one based on the plant feed conveyor rate over the calibration ranges. In spite of the visual similarity, based on the 95% confidence limit intervals these two curves are statistically different.

There is however another way of comparing these results. The standard errors of 1.94 and 1.05 for the slope and intercept, respectively, for the on-line calibration appears to be quite low considering the noise present on the data used in the

	Lab Calibration		On-line Calibration			
			Plant Feed Conveyor Rate		Collection Conveyor Rate	
	Value	95% Limit	Value	95% Limit	Value	95% Limit
Slope	29.81	35.27 24.36	51.90	55.82 47.97	45.95	49.76 42.15
Intercept	22.84	26.39 19.29	31.59	33.73 29.46	32.18	34.25 30.11

Table 2-1: Summary of calibration results.

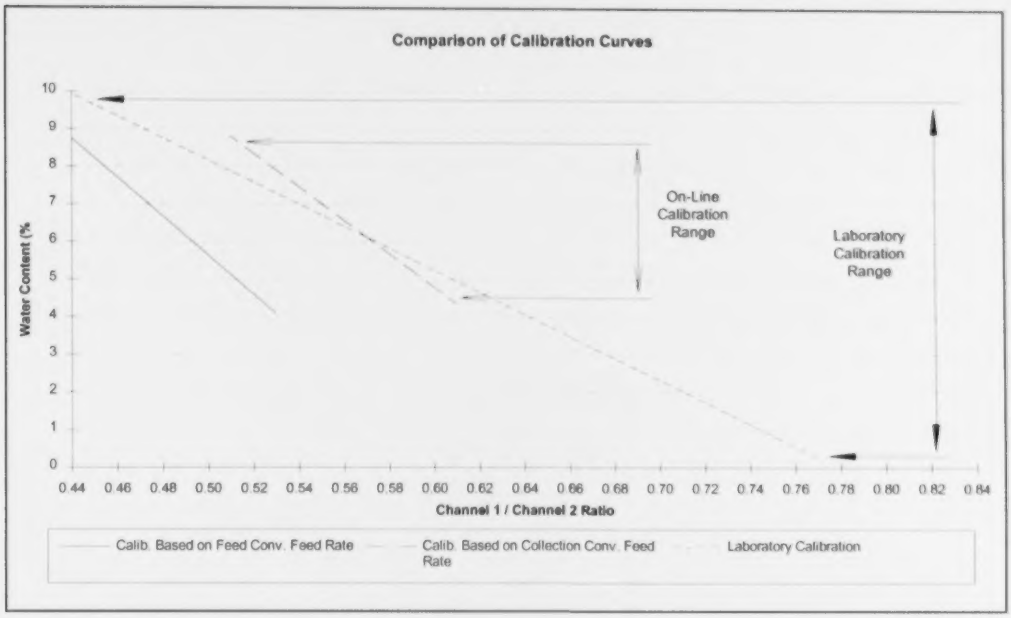

Figure 2-5: Comparison of the laboratory calibration curve and the two on-line calibration curves.

calibration. This is most likely due to the large number of data points for the regression. Since the on-line calibration is done using a density based model, it is impossible to distinguish between bitumen, with a density of $1.008\ \frac{kg}{L}$, and water at just under $1\ \frac{kg}{L}$. The accuracy of the bitumen measurement has been estimated at ±0.5% absolute. It is reasonable to assume that, since bitumen and water are measured on the same instrument using the same technique, their respective accuracies are of the same order. The way the on-line calibration was carried out, all the error is associated with the water measurement and the bitumen is assumed to be absolutely accurate. Therefore, what can be expected is that the on-line water calibration should give results that are within ±1% absolute of the true value. Over the valid range for the on-line calibration this holds true for the results obtained when the collection conveyor feed rate is used. At the upper end of the on-line calibration range the absolute difference is 0.85% and at the lower limit it is 0.81%, which is

within the anticipated ±1% absolute accuracy range.

The water content in the oil sand feed over the time period from noon on July 29th to midnight on August 1st is shown in Figure 2-6. As expected, it shows quite clearly that there is a good agreement between the laboratory calibration and the on-line calibration when the collection conveyor feed rate is used. The plant feed conveyor rate based calibration shows quite a large bias. Also, the average water content given by the latter calibration, of approximately 3%, is rather low when compared with what has been observed when fresh oil sand samples are analyzed analytically. This is another indication that the other two calibrations yield more consistent results.

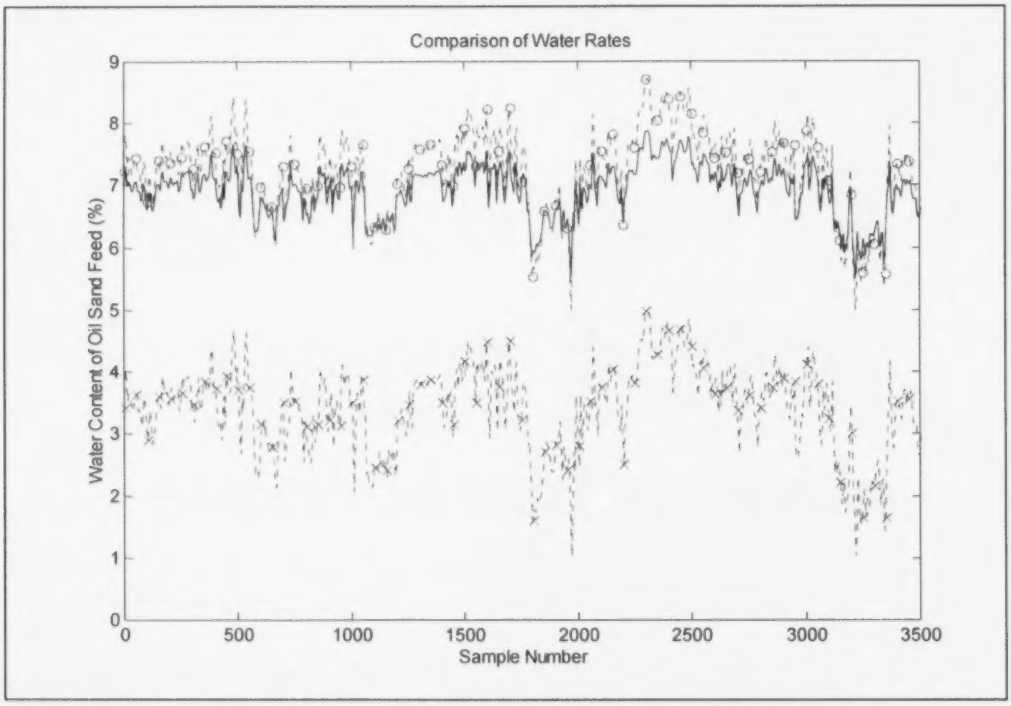

Figure 2-6: Comparison of water content measured using the laboratory calibration curve (—) and the two on-line calibration curves: based on plant feed conveyor rate (--x--) and collection conveyor rate (--o--).

2.8. Effect of uncertainty in the process measurements

Since each process measurement used in the on-line calibration does have different amounts of uncertainty associated with it, the effect of this uncertainty was evaluated. Whenever possible the variance chosen for this test was close to what one would expect from the type of sensor employed. The analysis examined the effect on the on-line water calibration one measurement at a time. No attempt was made to define an expected variance band that combined the uncertainty of all measurements. This analysis was carried out on the calibration which used the collection conveyor for the feed rate, since it is in better agreement with the laboratory based calibration. In most cases the variations used in this analysis are given as a percentage change in the sensor reading. Changing a sensor reading by -5% for example, is equivalent to saying that the sensor's output is biased high by 5%. Only sensor bias needs to be considered, since random fluctuations around the mean sensor reading are averaged out by the least squares calculation, regardless of their magnitude. The biases investigated are as follows:

- Oil sand feed rate measurement: the oil sand feed rate was varied by ±10% relative. In light of the discrepancy between the two redundant measurements for this quantity, this probably does not cover the complete range of variability.

- Reject density: for this part of the analysis, the effect of fixed reject densities of $2.25\,\frac{kg}{L}$ and $1.75\,\frac{kg}{L}$ was investigated.

- PSV slurry density measurement: this data was varied by ±2.5% relative.

- Steam condensation rate: the factor for steam condensation inside the

tumbler was changed from the nominal value of 0.75 by ±33.3% resulting in $\alpha = 1.0$ and $\alpha = 0.5$.

- Flood water rate measurement: the flood water flow readings were changed by ±5%, relative, from their recorded values.

- Tumbler water measurement: these values were varied by ±5% relative.

- Bitumen concentration: these readings were changed by ±0.5% absolute, which represents the known accuracy range for this measurement.

A summary of the limits established by the sensitivity analysis results is given in Table 2-2. Figures 2-7 through 2-13 are graphical representations of the result of the sensitivity analysis over the calibration range and compare them to the laboratory calibration curve. As can be seen in these graphs, in every case the on-line calibration is only close to the laboratory based calibration over the narrow range over which the on-line calibration was carried out. Extrapolation beyond this range results in significant discrepancies between the two.

Feed rate variations have the largest impact on the calibration constants only because of the magnitude of their uncertainty. If the confidence in the feed rate measurement would be higher, the impact would be greatly reduced. The calibration method is most sensitive to the slurry density measurement. This in not entirely unexpected, since the method is based on this parameter. Fortunately, quantities that are impossible to measure, such as reject density and steam condensation rate inside the tumbler, have a minimal impact on the results.

Measurement	X_1	Change in X_1	X_2	Change in X_2
Reference	31.59		44.66	
Feed rate +10% (relative)	32.01	1.32%	40.71	-8.85%
Feed rate -10% (relative)	31.08	-1.61%	49.49	10.81%
Reject density fixed at 2.65 $\frac{kg}{L}$	30.94	-2.08%	44.34	-0.71%
Reject density fixed at 2.00 $\frac{kg}{L}$	32.44	2.67%	45.07	0.91%
Slurry density +2.5% (relative)	29.80	-5.67%	47.99	7.46%
Slurry density -2.5% (relative)	33.47	5.96%	41.16	-7.84%
Steam condensation +33% ($\alpha = 1.0$)	31.25	-1.10%	44.75	0.21%
Steam condensation -33% ($\alpha = 0.5$)	31.94	1.10%	44.56	-0.21%
Flood water flow +5% (relative)	31.97	1.20%	46.58	4.30%
Flood water flow -5% (relative)	31.22	-1.20%	42.74	-4.30%
Tumbler water flow +5% (relative)	31.10	-1.56%	44.89	0.53%
Tumbler water flow -5% (relative)	32.08	1.53%	44.42	-0.53%
Bitumen content +0.5% (absolute)	31.12	-1.49%	44.66	0.00%
Bitumen content -0.5% (absolute)	32.07	1.49%	44.66	0.00%

Table 2-2: Summary of sensitivity analysis results.

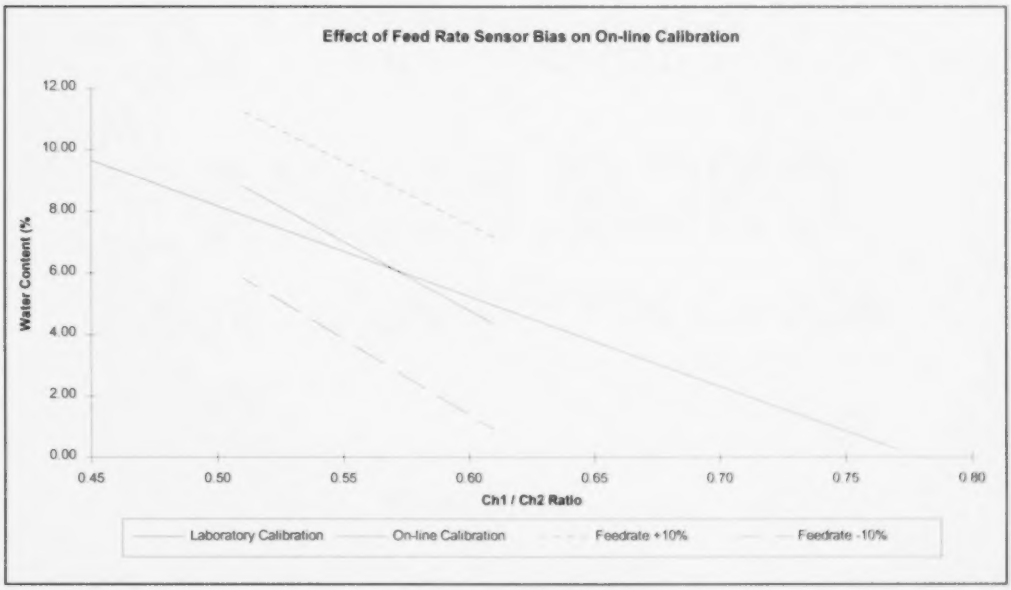

Figure 2-7: Effect of a 10% relative bias in the feed rate on the on-line calibration curve.

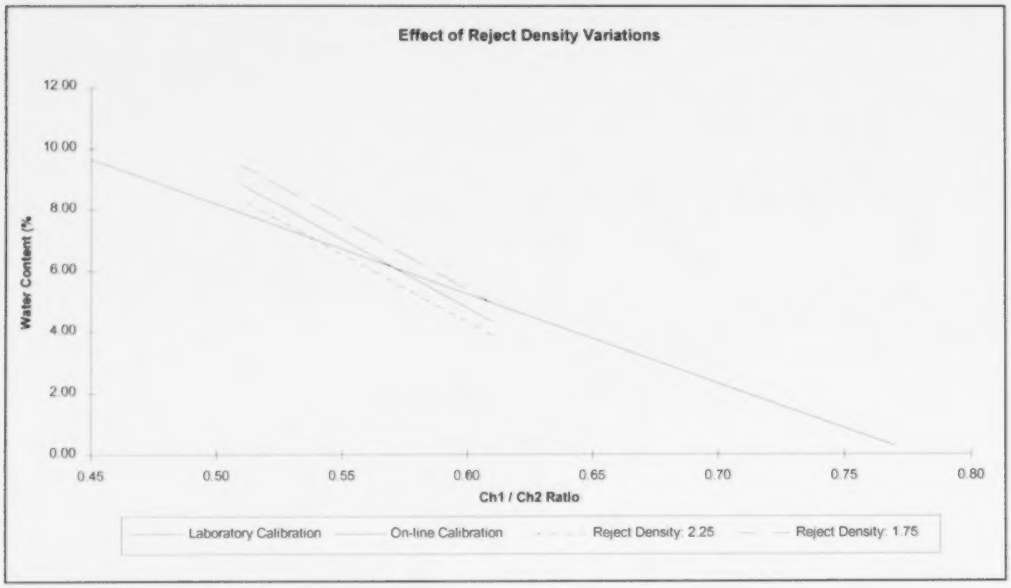

Figure 2-8: Effect of errors in the assumed reject density on the on-line calibration curve.

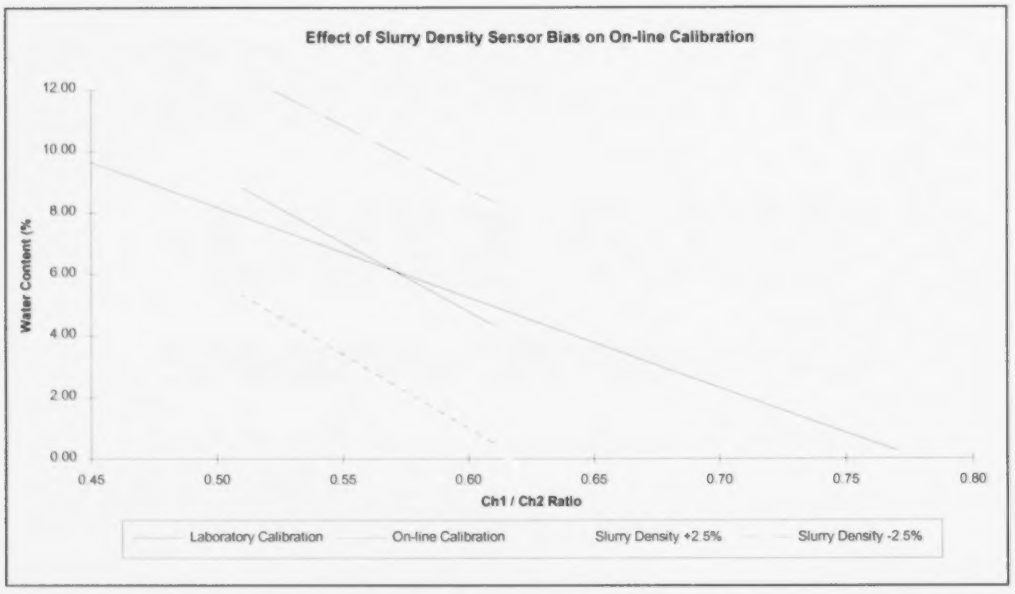

Figure 2-9: Effect of a 2.5% relative bias in the PSV slurry density measurement on the on-line calibration curve.

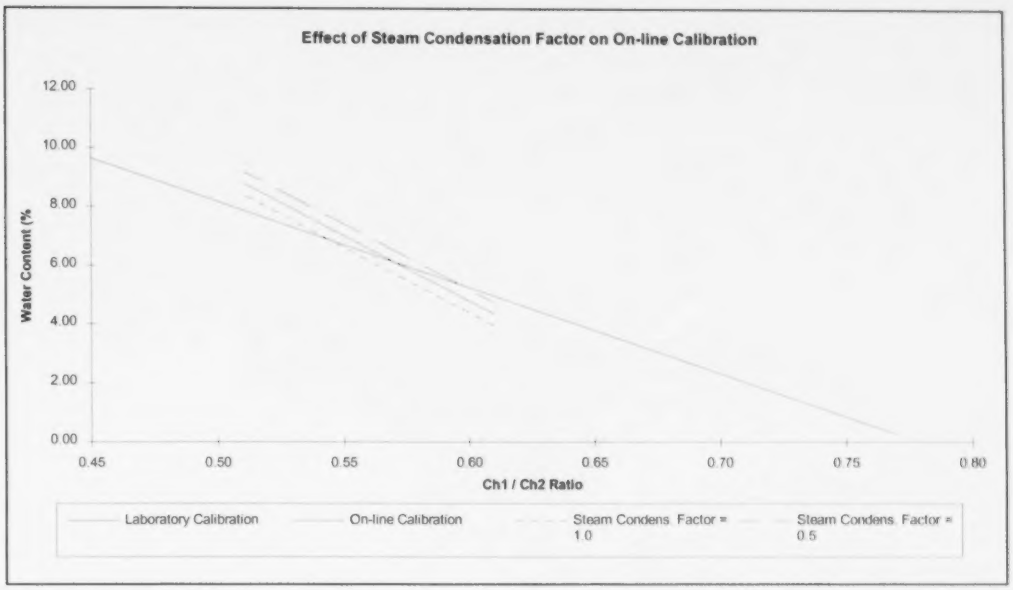

Figure 2-10: Effect of changes in the steam condensation factor inside the tumbler on the on-line calibration curve.

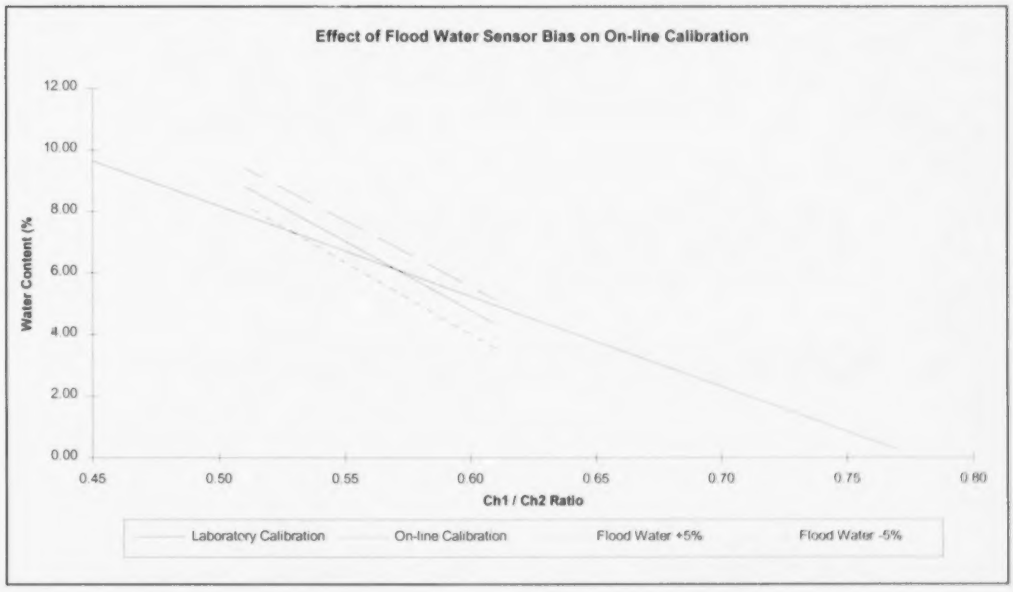

Figure 2-11: Effect of a 5% bias in the flood water flow rate measurement on the on-line calibration curve.

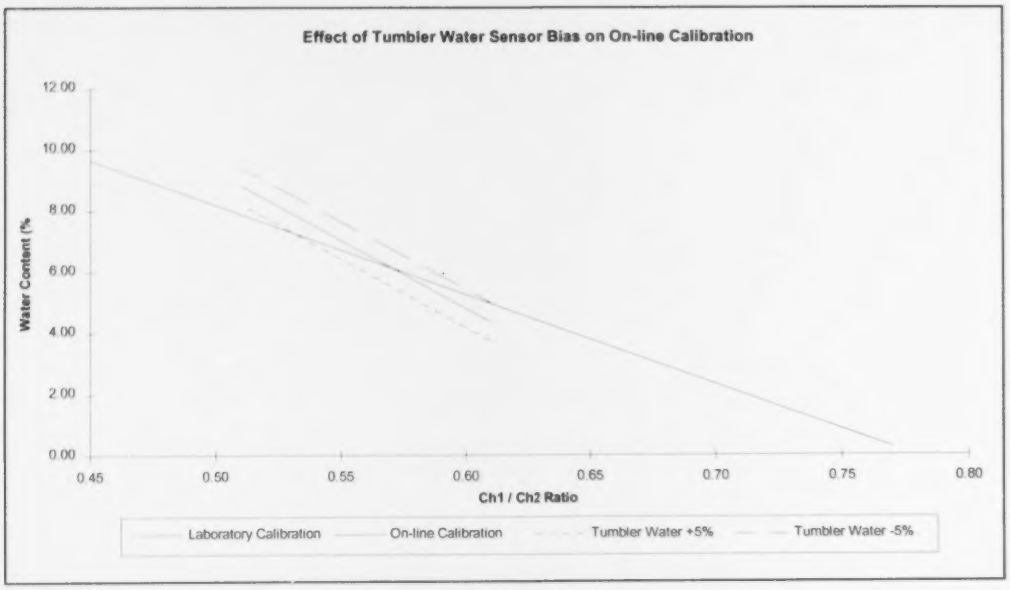

Figure 2-12: Effect of a 5% bias in the tumbler water flow rate measurement on the on-line calibration curve.

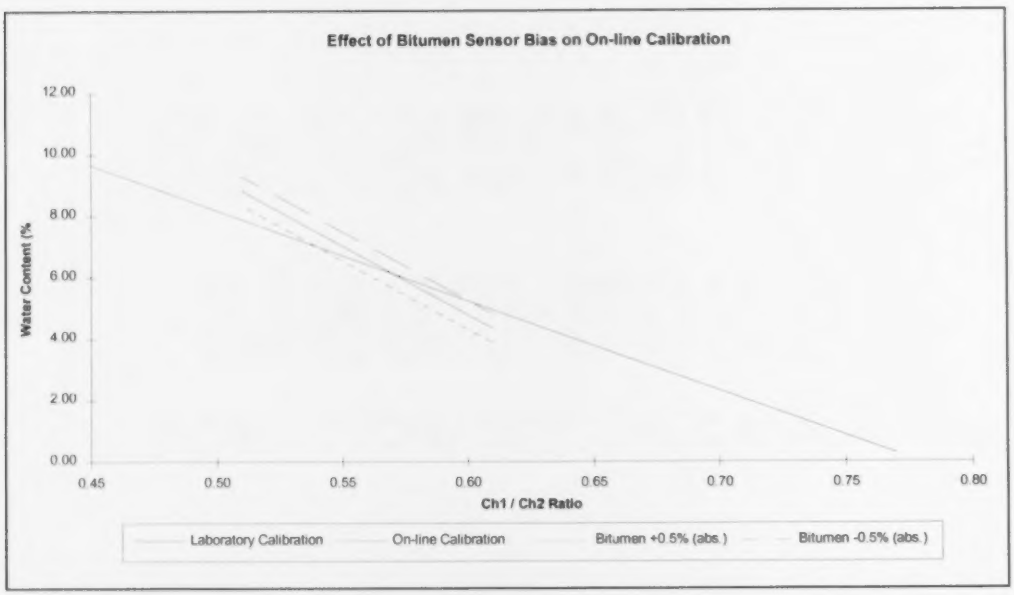

Figure 2-13: Effect of a 0.5% absolute bias in the bitumen content determination on the on-line calibration curve.

2.9. Feed rate evaluation

The rather large differences between the feed rate measured at the collection conveyor and the one measured at the plant feed conveyor are shown in Figure 2-14. The feed rate in both cases is determined using a nuclear weigh scale. As has been pointed out by a scientist at Syncrude's Research department[1], the measurement of feed rate from these units is heavily influenced by the deposition shape of the material on the conveyor belt. Through simulations he has shown that, using the exact same mass of oil sand and changing only its distribution across the belt, the feed rate reading from this instrument changes by up to 20%. The discrepancy in the observations presented in Figure 2-14 are close to this value. Since the two measurements are taken on separate conveyor belts that are at right angles to each

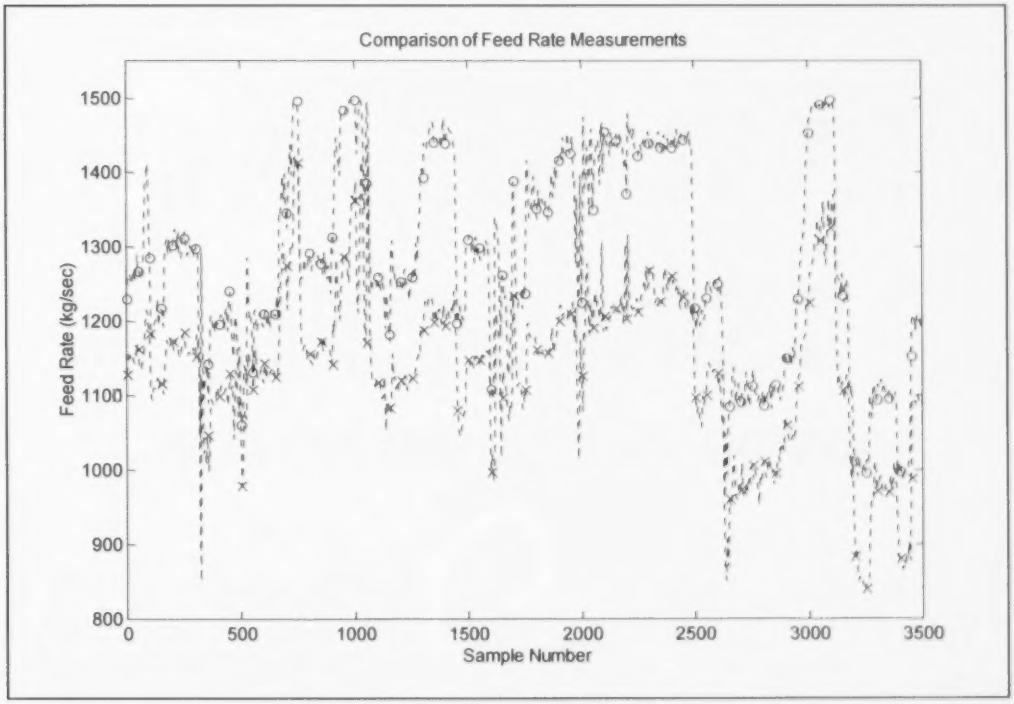

Figure 2-14: Comparison of plant feed conveyor (--x--) and collection conveyor (--o--) feed rates.

other and have a transfer point between them, it is reasonable to expect that the way the oil sand is distributed on the two belts is quite different. An additional source of error between these two measurements could be instrument bias, since these units are quite difficult to calibrate. Figure 2-15 shows a plot of the plant feed conveyor feed rate versus the collection conveyor feed rate. The expected relationship between these two measurements is indicated by the dashed line in this figure. The fact that all but one of the data points are above this theoretical line points to the presence of a bias between the two instruments. However, simple re-calibration can not address the measurement errors due to the changing oil sand profile on the belts. On the other hand, since the general trend in the data is parallel to the expected line, it can be concluded that the oil sand distribution on the two belts is similar.

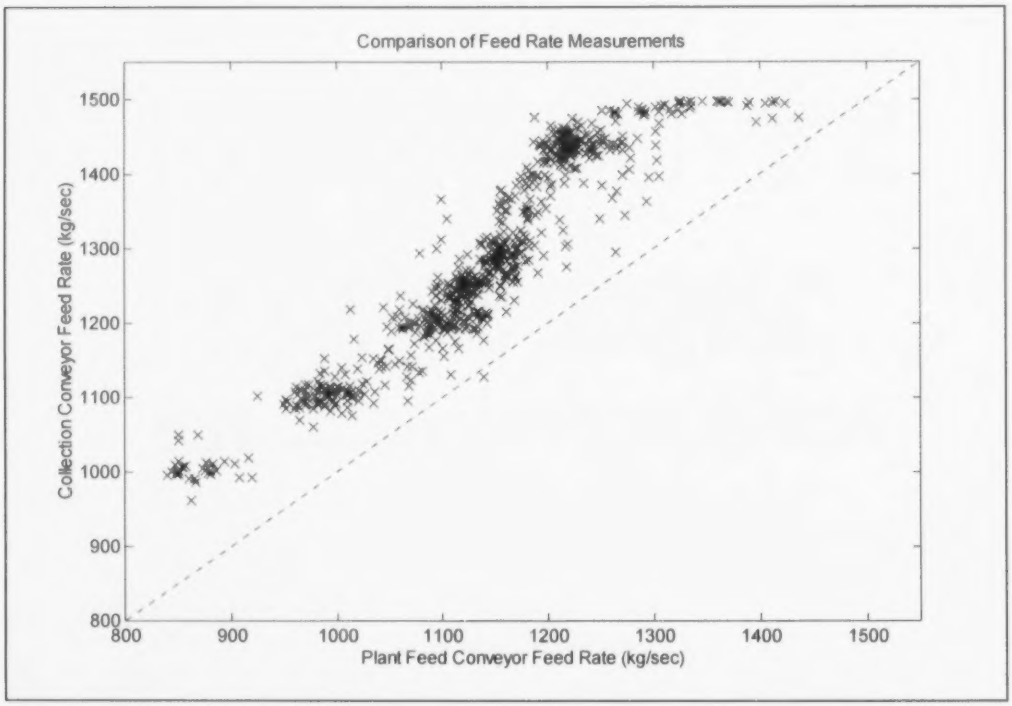

Figure 2-15: Plot of the plant feed conveyor feed rate versus the collection conveyor feed rate.

Method for determining feed rate	Total Feed (kg)	Average Feed Rate $\left(\frac{kg}{sec}\right)$
Collection conveyor measurement	4 530 665	1 281
Plant feed conveyor measurement	4 019 259	1 137
Model	4 427 482	1 252

Table 2-3: Comparison of different oil sand feed rate determination methods.

It is impossible to devise a truly independent method of determining which of the two feed rates is closer to the actual one. But, if one accepts the water calibration obtained in the laboratory for the Oil Sand Monitor, then the model equations used for the on-line calibration can be rearranged to calculate a feed rate, which is given by:

$$FR = \frac{\rho_S \rho_B \left(\rho_R \left\{ \left[\rho_{Slurry} - \rho_W^{Slurry} \right] \left[\rho_W^T (TW) + \alpha (SR) + \rho_W^F (FW) \right] + \rho_W^{Slurry} m_R \right\} - \rho_{Slurry} \rho_W^{Slurry} m_R \right)}{\rho_R \left\{ \rho_{Slurry} \rho_W^{Slurry} \left[\rho_B (B) + \rho_B (W) - \rho_B - \rho_S (B) \right] + \rho_S \rho_B \left[\rho_W^{Slurry} - \rho_{Slurry} (W) \right] \right\}} \tag{2.6}$$

The assumptions made for α, ρ_W^T, ρ_W^F, ρ_W^{Slurry}, ρ_B, ρ_S and ρ_R stated earlier were not changed for this investigation. A comparison of the two measured and the calculated feed rates is presented in Figure 2-16. The calculated feed rate, for the most part, is close to the one measured at the collection conveyor, but does have significant deviations at times from either measurement. Table 2-3 summarizes the total amount of feed over the 59 hour interval used here, as well as the average feed rate to the plant over this time period. The large discrepancy between the two measurements has a serious impact not only on both the production and recovery calculations for the plant, but also on the tumbler control loops.

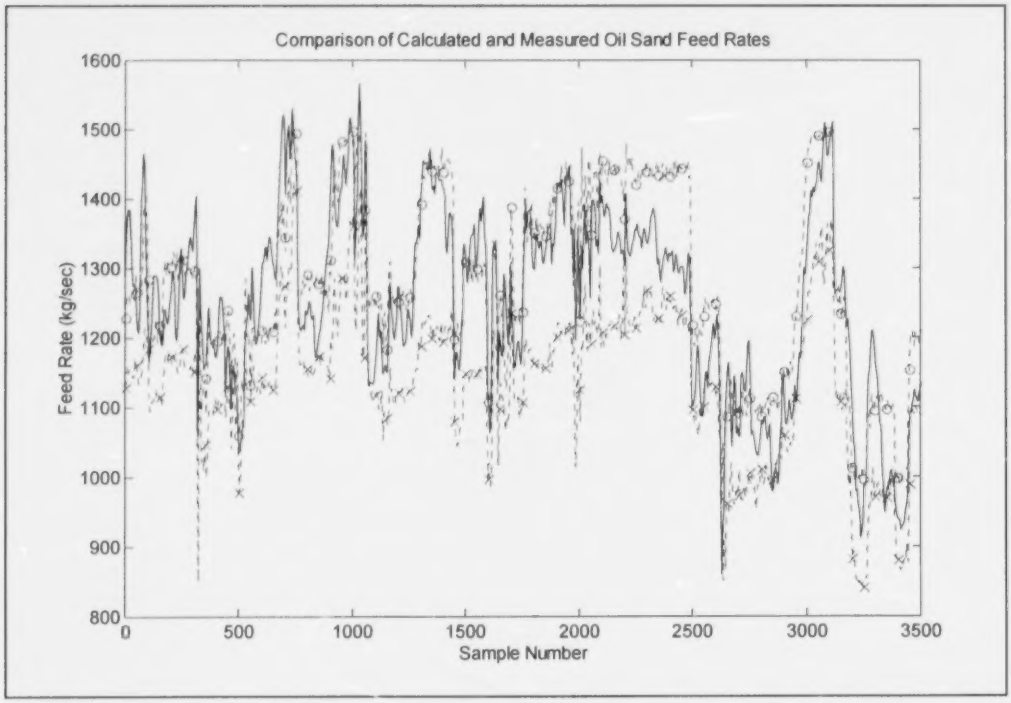

Figure 2-16: Comparison of the calculated feed rate (—) and the two measured feed rates: plant feed conveyor (--x--), collection conveyor (--o--).

2.10. Conclusion

For the first time since the initial installation of the Oil Sand Monitor in 1983 the investigation presented here has proven that, in spite of the significant difference in conditions between the laboratory calibration and on-line operation, the results obtained from this instrument reflect the true oil sand composition as the material moves past on the conveyor belt.

Using a model of the process to calibrate an on-line sensor is theoretically possible, but the accuracy requirements for all measurements that are used in the model are quite severe and probably can not be met in a majority of the cases. The observed deviations between the laboratory and the collection conveyor feed rate

based on-line calibrations are within the expected limits over the valid calibration range of the on-line calibration. However, it is quite evident that outside the narrow on-line calibration range this no longer holds true. This leads to another concern with model based on-line calibration methods. Since stable plant operation is essential to collecting a good data set for the calibration calculations, it is unlikely that this coincides with large fluctuations in the property of interest, resulting in a calibration over a very narrow range. Extrapolation, usually linear, beyond the calibration range raises serious questions about the validity of the readings obtained. Random variability in the sensor readings does not create a problem for on-line calibration, provided the data set used is sufficiently large. Biased sensor readings on the other hand do affect the calibration constants, often significantly, as was seen in the sensitivity analysis. A sensitivity analysis is a crucial part of the on-line calibration process since it can be used to determine if the desired accuracy for the calibration can be achieved. For example, if it shows that a certain parameter must be known to an accuracy of ±0.1%, but can at best be measured to ±1%, then an on-line calibration is not possible unless the accuracy expectations for the calibration are lowered. The conclusion therefore is that a careful laboratory calibration is still the best way of calibrating an instrument. However, if laboratory conditions can not approximate the service conditions in the plant, a study similar to this one should be carried out to confirm that this difference in conditions does not affect the measurement. If the performance evaluation shows that the laboratory calibration does not yield satisfactory results, a model based on-line calibration can be carried out, bearing in mind all the limitations stated above.

The fact that the feed rate is measured independently in two locations is fortunate, considering the measurement problems with the nuclear weigh scales. However, changing the way this measurement is carried out, to make it independent

of the oil sand deposition shape would have a positive impact far beyond the on-line calibration and performance evaluation of the Oil Sand Monitor. Nevertheless, even with the feed rate measurements available today, several improvements to the process are possible. As shown in the section on feed rate evaluation, it is possible to use the model developed here to calculate a feed rate, the reliability of which is most likely greater than the reliability of the actual measurements as they are currently carried out. This calculated feed rate can be used in determining recovery and throughput estimates for the extraction plant. Unfortunately, since the calculated feed rate is based on the PSV slurry density, which is measured after the tumbler, it is impossible to use this value for controlling tumbler water and caustic addition. Currently these control loops use only the plant feed conveyor rate as their input. It would be equally possible to use the collection conveyor rate for this purpose. A control strategy could then be devised where the calculated feed rate could be used to select which of the two feed rates should govern the tumbler control loops. This could be updated at maybe 10 or 15 minute intervals depending on which of the two rate measurements was closer to the calculated feed rate over the past interval. While this does not give optimal control of the tumbler, it would represent an improvement over the current control strategy.

2.11. References

[1] Dougan, P.D., "An Analysis of Nuclear Weighing Systems in Extraction." Confidential internal Syncrude Research report.

3. Spectral Data Analysis and Filtering

Data collected from operating industrial plants often contains a substantial noise component. Noise in a data stream are observed variations that are independent of the process dynamics. These variations can be either random or systematic. Often the source of the noise is inherent to the sensors employed to obtain the data or to the characteristics of the process equipment and can not be eliminated. Careful examination of the noise component of a signal will lead to better process models and better controller designs.

When noisy data is used for process identification, the noise will increase the uncertainty in the models that are developed by obscuring the true dynamics of the process and in some cases introducing artificial dynamics. Using this data 'as is' for process control will degrade the controller performance and can propagate artificial dynamics to other parts of the process.

Filtering of data for identification and control is an effective way of attenuating unwanted components from the data. At this stage, a good understanding of the process is essential in defining the parameters of the filter. Unfortunately, a simple time series plot of the data is often not sufficient for filter design. Spectral analysis of the data is helpful in identifying the frequency components of the signal on which the filter parameters can be selected.

In the remainder of this chapter, a case study involving the oil sand feed rate signal for Syncrude's extraction plant will be presented.

3.1. Spectral analysis of feed rate signal

The data used for this analysis consists of 2820 values collected over a 47

hour period at the rate of one data point per minute. A time series plot of this data is presented in Figure 3-1. In order to characterize the nature of the noise present in this signal, a fast Fourier transform (FFT) was performed on the data. The frequency spectrum resulting from this analysis is given in Figure 3-2. Two distinct features were found in this spectrum: a very sharp peak at a period of 4 minutes and a much broader peak at 720 minutes.

The 720 minute period corresponds to the 12 hour shift for the plant operators. It can be interpreted in two ways. One interpretation is that each operator sets the feed rate for the duration of the shift at a different level. Another explanation could be that during the shift change period, the feed rate is temporarily adjusted to facilitate in some way the transition from one crew to the next. Further investigation is required to determine which of the two scenarios is the most likely.

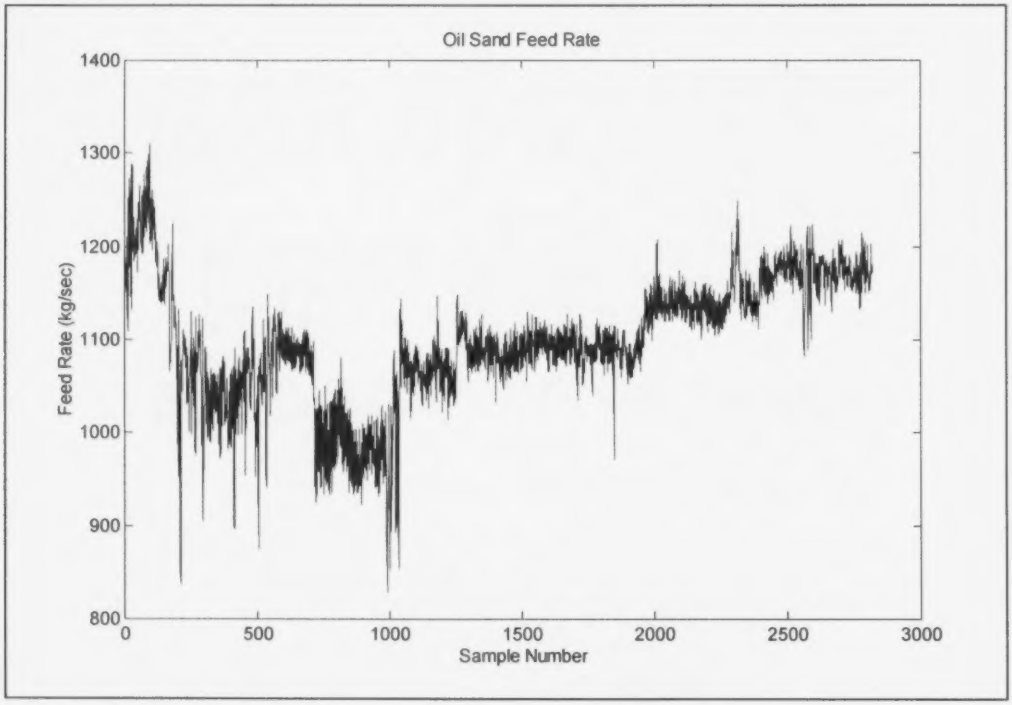

Figure 3-1: Time series plot of oil sand feed rate.

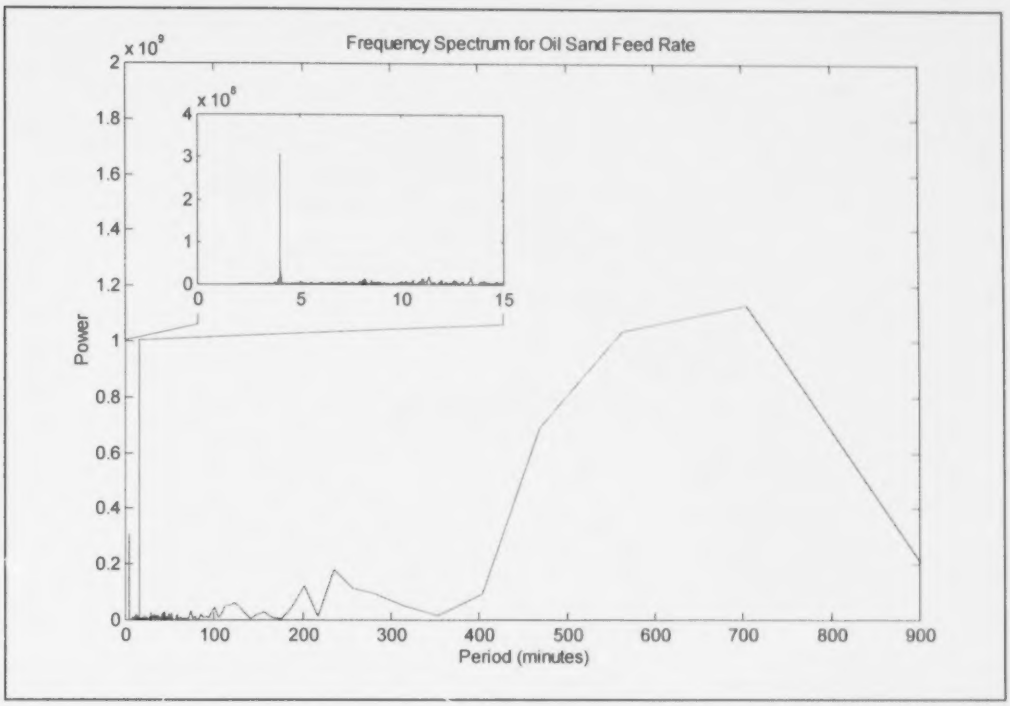

Figure 3-2: Frequency spectrum of oil sand feed rate.

The feature that has the largest impact on the operation of the tumbler water control loop is the sharp spike at a period of 4 minutes. This corresponds to the time required for the plant feed conveyor belt to make one complete loop. The source of this disturbance was found to be the measurement system itself. Determination of feed rate is done by using a nuclear weigh scale. A radiation source emits a narrowly focused beam of gamma radiation down onto the plant feed conveyor belt. The detector mounted just below the belt measures the amount of radiation that passes through the oil sand and the belt itself. A decrease in the radiation measured by the detector is interpreted as an increase in the amount of material on the belt, and hence a higher feed rate. Conversely, an increase in the observed radiation corresponds to a decrease in the feed rate. Normally the belt attenuates this radiation by a constant amount. However, when the section of the belt where the two ends are spliced together passes in front of the detector, the attenuation due to the belt is increased.

The reason for this lies in the belt construction and in the way the belt ends are spliced together. The belt has a reinforcing core made up of a substantial number of steel cables spaced evenly across its width. In order to connect the two ends of the belt together, the steel cables from one end of the belt are laid alongside the ones from the other end and the top and bottom rubber layers are sealed together. This in effect doubles the number of cables over the distance of the splice section. The decrease in the radiation seen by the detector due to the increased attenuation caused by the belt splice is interpreted as an increase in feed rate. One impact of this is that the feed rate to the plant will have slight bias. A more significant impact, however, is the effect on control loops that use this signal as one of their inputs.

3.2. Downstream effects

One of the control loops on which the oil sand feed rate has a direct impact is the tumbler water control circuit shown in Figure 3-3. The feed rate signal is fed to a ratio controller that maintains a constant oil sand to tumbler water ratio. The output from the controller actuates a flow control valve on the water line. Feedback to the controller is via a flow meter on the same line. A time series plot of both the controller output and the flow meter output is given in Figures 3-4 and 3-5, respectively. Spectral analysis of the two signals confirms that the 4 minute spike passes all the way through the system as shown in Figures 3-6 and 3-7. The fact that it can be seen in the output from the flow meter confirms that the flow controller actually reacts to the spike. The effect on the process is that the actual feed rate to water ratio does not remain constant at the desired level. What, if any, impact this has on the process is difficult to assess and may in practice not be significant. Another effect may be an insignificant amount of increase in the hot water consumption. The most significant impact, from a practical point of view, is the added movement in the

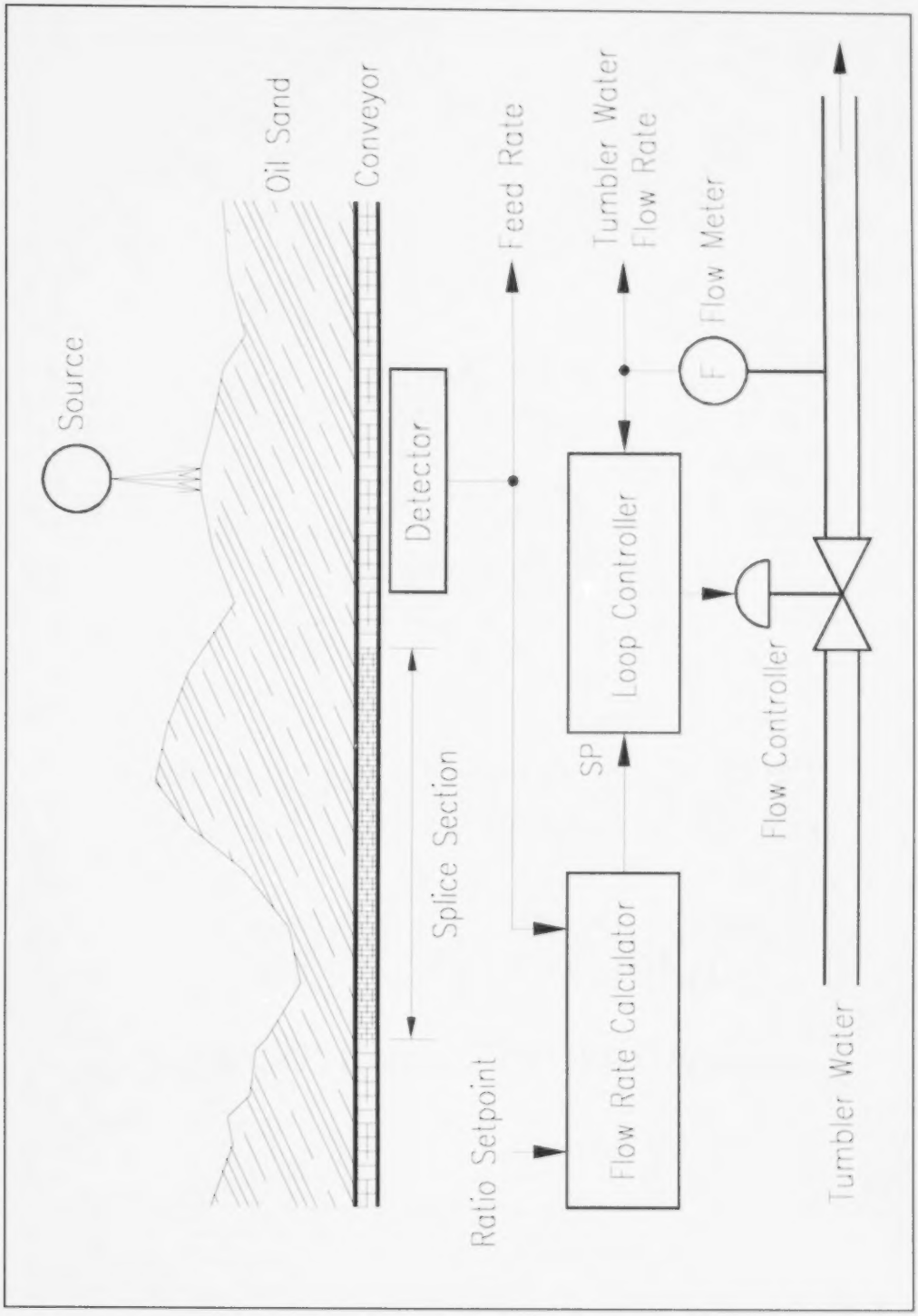

Figure 3-3: Tumbler water control loop schematic.

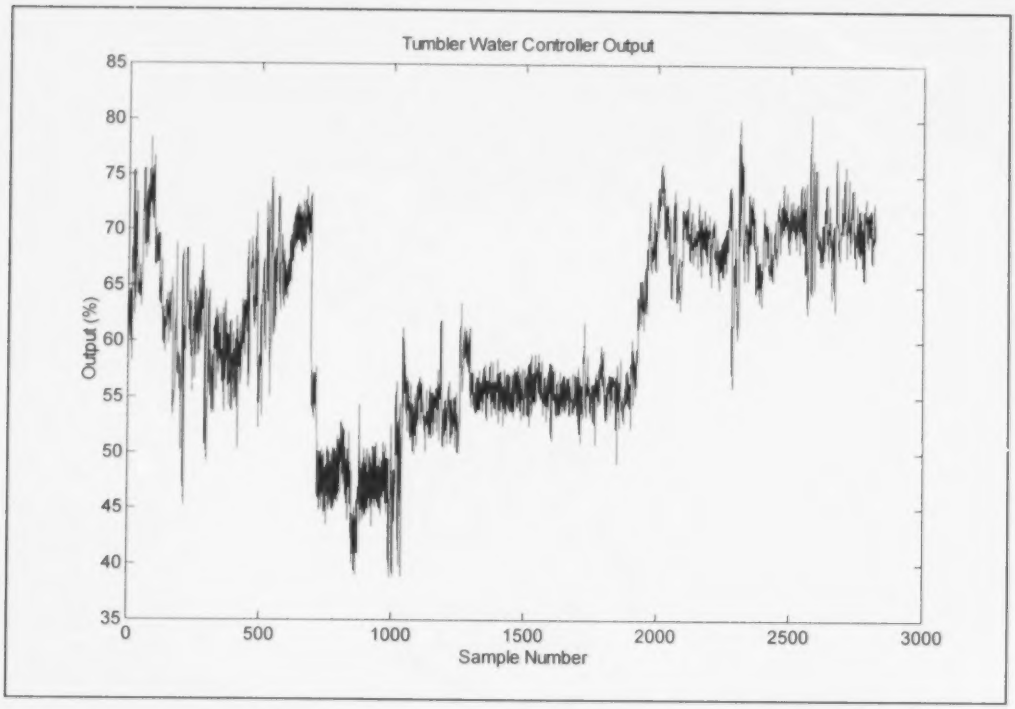

Figure 3-4: Time series plot of tumbler water controller output.

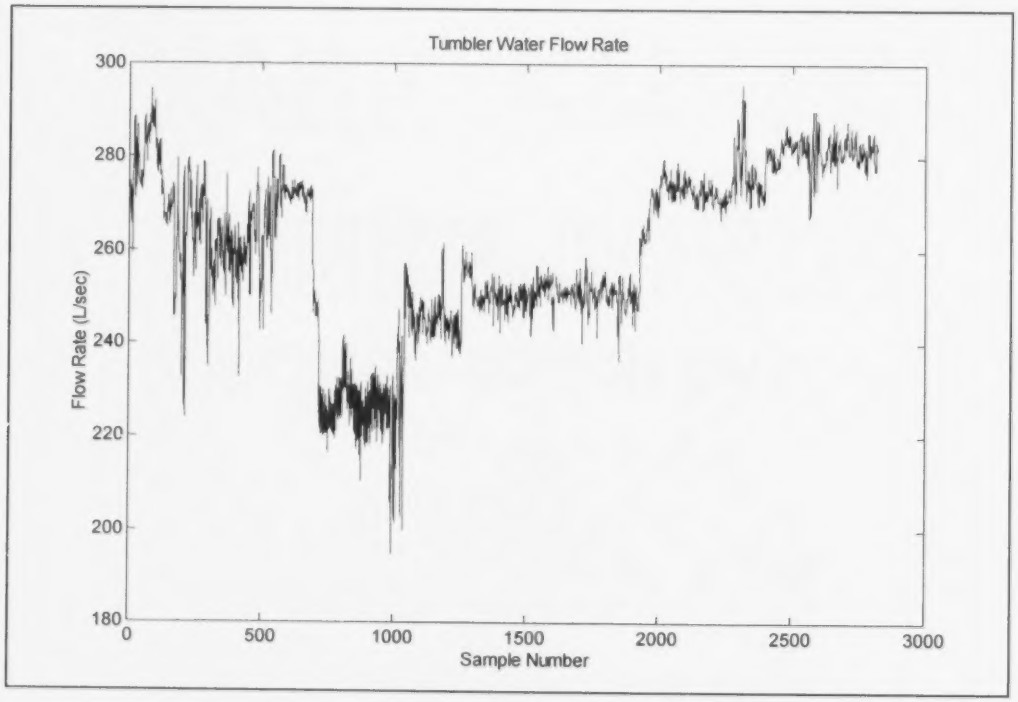

Figure 3-5: Time series plot of tumbler water flow.

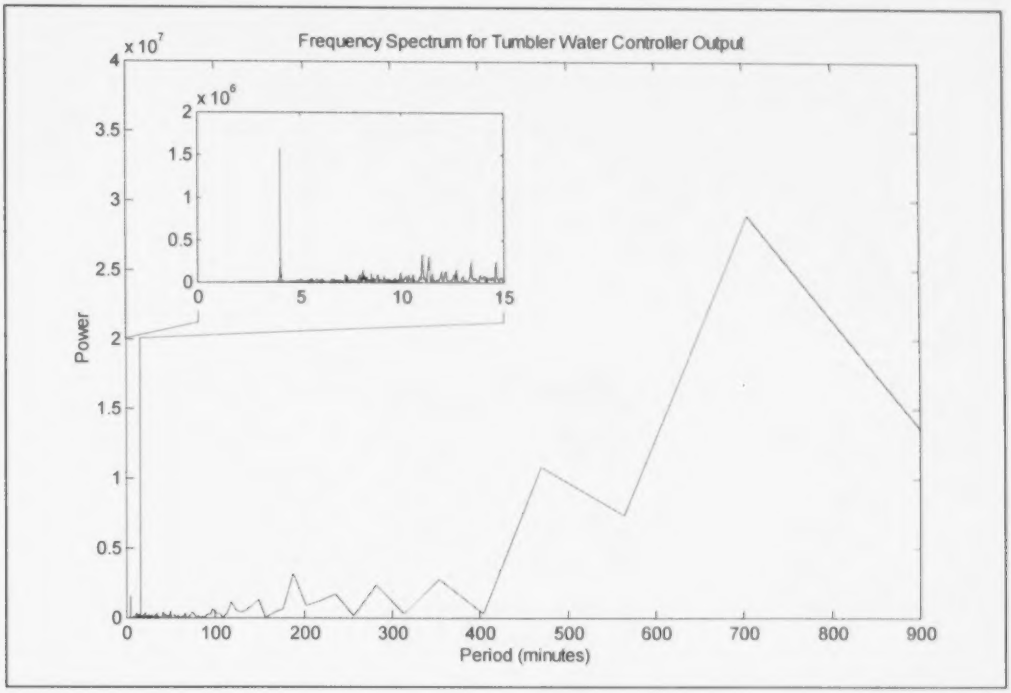

Figure 3-6: Frequency spectrum of tumbler water controller output.

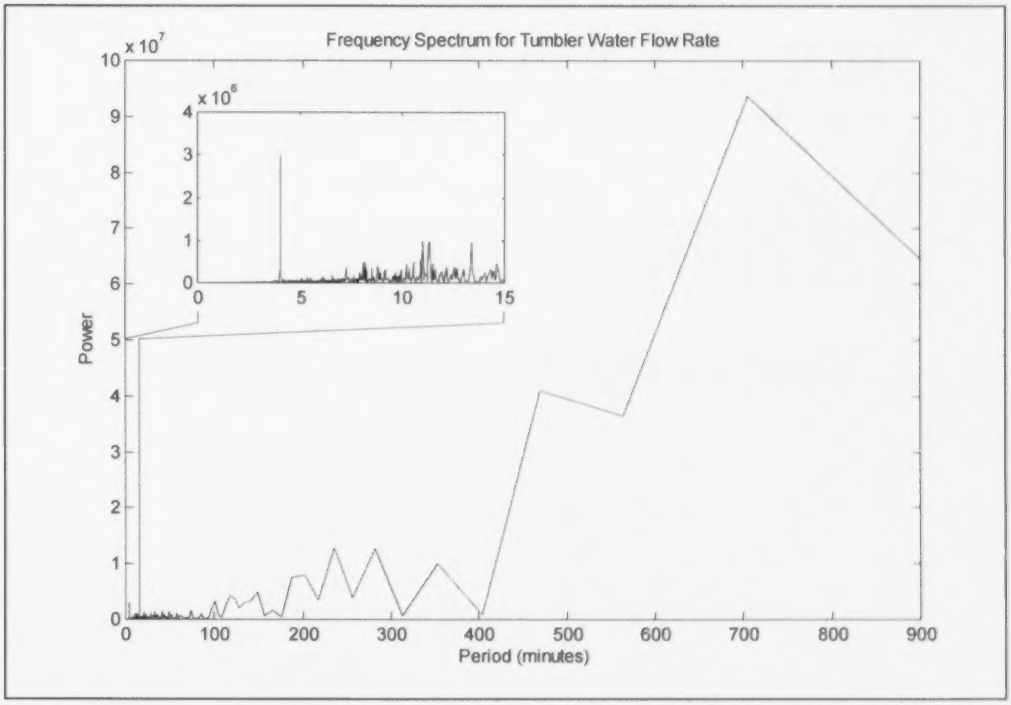

Figure 3-7: Frequency spectrum of tumbler water flow.

flow control valve and its actuator, causing added wear of the mechanical components.

The tumbler water controller output signal and the associated flow rate also exhibit a significant peak at 720 minutes. This period corresponds to the twelve hour operating shifts in the plant. This particular disturbance is much more pronounced in the tumbler water addition data in the feed rate and probably indicates differences in the ratio set point between different operators. Unfortunately this can not be confirmed, since the ratio controller set point is not logged in the plant data base.

3.3. Feed rate data filter design

As identified above, applying a filter to the oil sand feed rate should result in an improved tumbler water control loop performance. The main objective for the filter is the removal of the spurious spike at a period of 4 minutes in the frequency spectrum. The filter chosen for this application was a recursive infinite impulse response (IIR) type low pass filter. Several filters with a cutoff period of 5 minutes and an order ranging from one to eight were designed and their performance was compared (Figure 3-8). The frequency units (F) on the abscissa in Figure 3-8 is relative to the sampling period of the signal ($P_{sampling}$). Table 3-1 summarizes the significant points on the frequency unit scale. Frequency units can be converted to the period (P) by the following equation:

$$P = \frac{2P_{sampling}}{F} \tag{3.1}$$

As can be seen in Figure 3-8, low order filters are not capable of attenuating the signal at 0.5 frequency units significantly. High order filters on the other hand remove the offending frequency very well, but exhibit a significant overshoot at 0.35

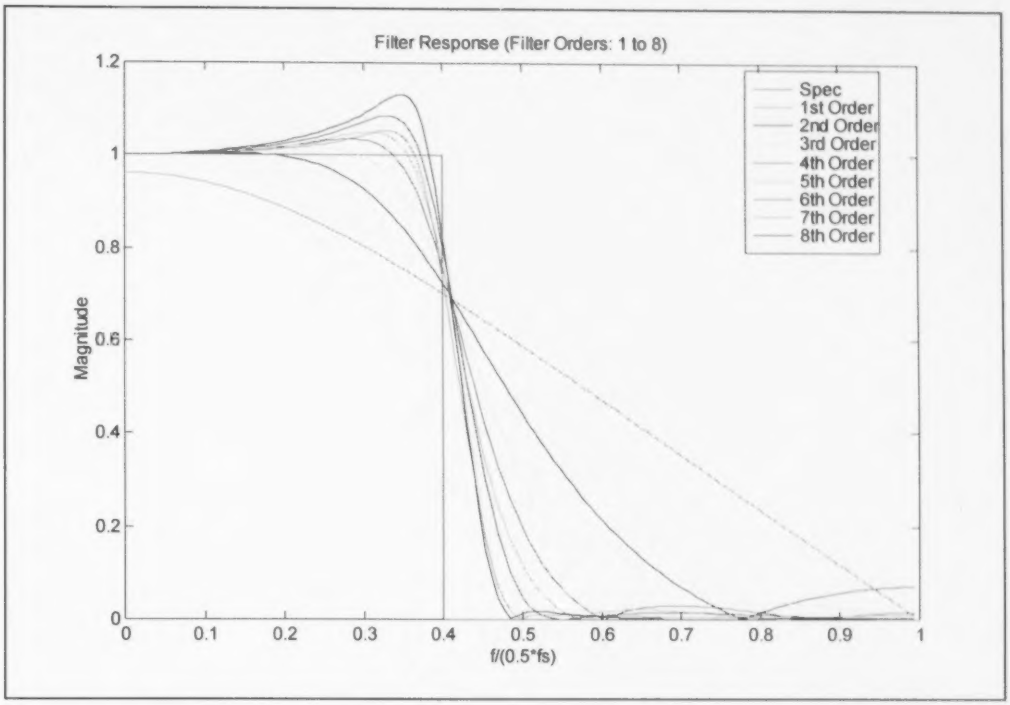

Figure 3-8: Filter performance evaluation.

frequency units. A fifth-order filter was chosen since it represented a good compromise between these two extremes. Realizing that it is impossible to achieve the filter specification as outlined in Figure 3-8 (100% of the signal with periods of 5 minutes and longer is passed and all of the signal with periods of less than 5 minutes is removed) several fifth-order order filter designs with different frequency response specifications were evaluated. The one selected is shown in Figure 3-9. The frequency design points for this filter are 0.3333 (6 minutes) for the upper limit of the

Frequency Unit	Period (minutes)
0.0	∞
0.4	5
0.5	4
1.0	2

Table 3-1: Significant frequency unit points.

roll off and 0.4444 (4.5 minutes) for the lower one. It attenuates the undesired spike with a 4 minute period by more than 90% and exhibits virtually no overshoot. The equation for this filter is given by:

$$y_n = 0.0631x_n + 0.1485x_{n-1} + 0.2182x_{n-2} + 0.1962x_{n-3} + 0.1186x_{n-4} + 0.0308x_{n-5} + 1.0222y_{n-1} - 1.1235y_{n-2} + 0.5068y_{n-3} - 0.2210y_{n-4} + 0.0407y_{n-5} \quad (3.2)$$

where: y_n = current filter output
y_{n-i} = filter output at the i^{th} previous time step
x_n = current filter input
x_{n-i} = filter input at the i^{th} previous time step

The time series plot and the frequency spectrum plot for the filtered feed rate are shown in Figures 3-10 and 3-11, respectively. As can be seen from the frequency spectrum, the spike at 4 minutes has been removed without affecting the frequency

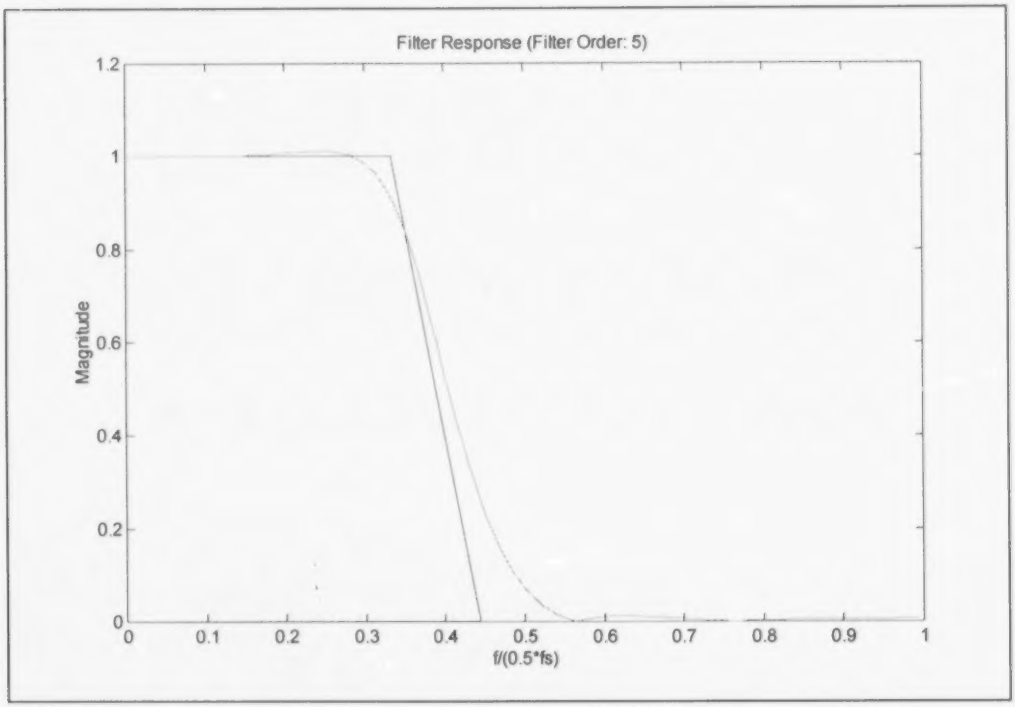

Figure 3-9: Response of filter used for filtering the oil sand feed rate.

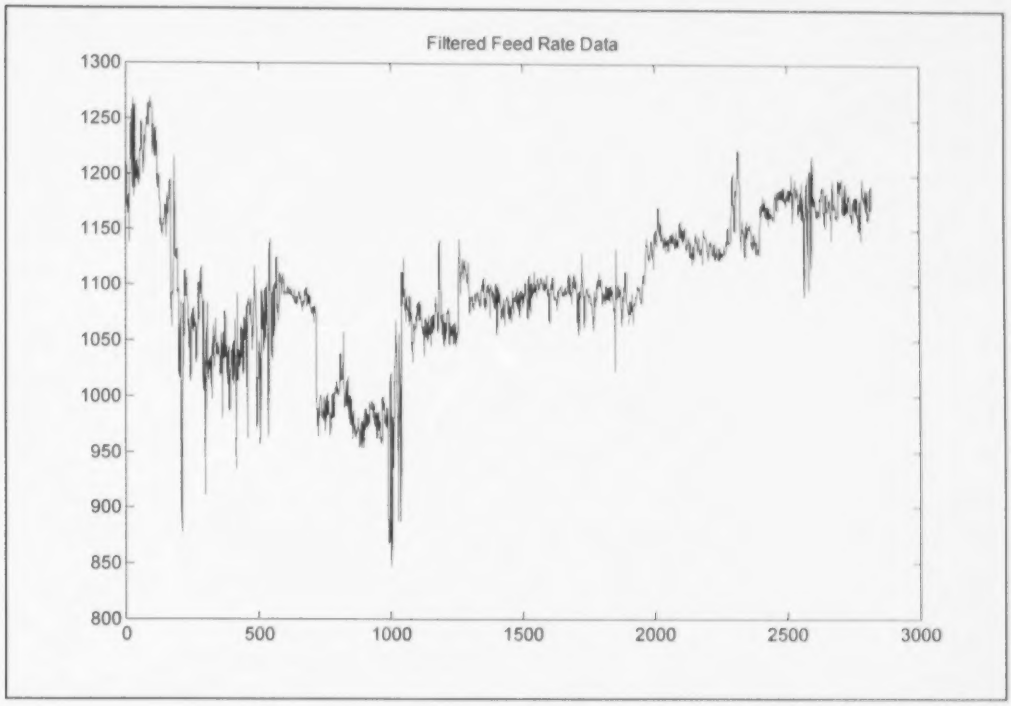

Figure 3-10: Time series plot of filtered oil sand feed rate.

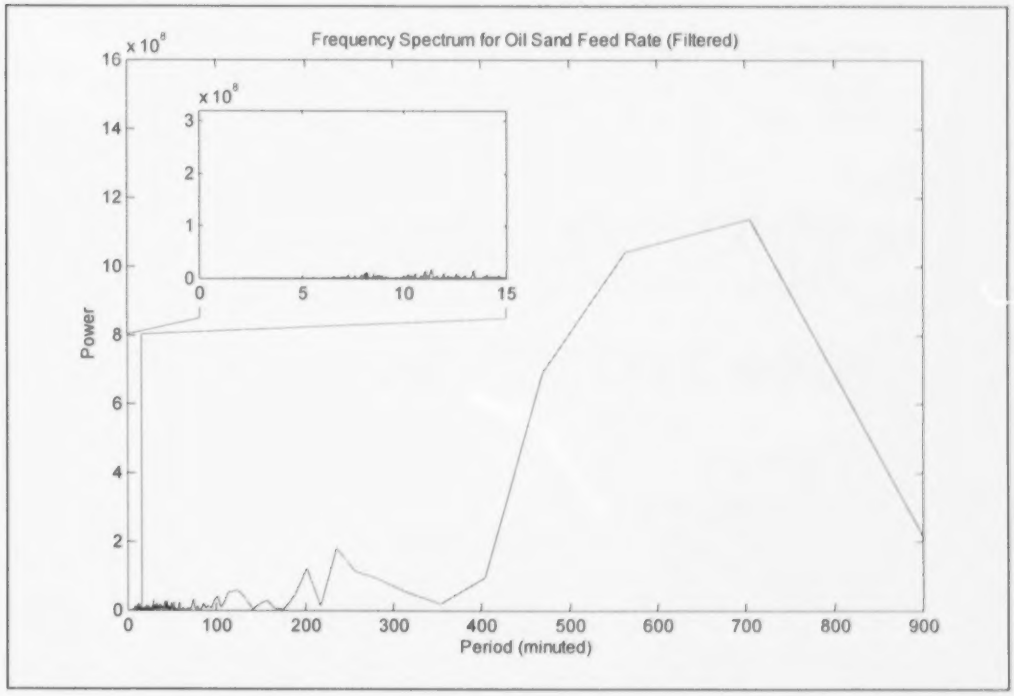

Figure 3-11: Frequency spectrum of filtered oil sand feed rate.

response of the rest of the signal. Furthermore, comparing Figures 3-1 and 3-10, it is evident that the feed rate signal has been smoothed considerably without removing or distorting any important feed rate fluctuations.

3.4. Simulation results

In order to evaluate the effect of the filtered feed rate on the amount of water added to the tumbler and on the control effort for the control loop, a simulation was carried out. The results for the simulation are given in Table 3-2. To verify the accuracy of the simulation, it was also carried out on the raw (unfiltered) feed rate data, establishing a benchmark for the simulation.

As can be seen from these results, the simulation of water addition using the raw feed rate data agrees well with the actual volume of water added. Using the filtered feed data, this volume did not change significantly. However, a dramatic improvement in the control effort was observed. The control effort was defined as the sum of all incremental control moves carried out over the simulation interval:

$$CE = \sum_{i=2}^{N} |x_i - x_{i-1}| \tag{3.3}$$

The percent control effort for the simulation with the filtered feed rate was referenced to the raw feed data case as follows:

	Raw Feed Rate	Filtered Feed Rate
Total volume of water added (actual)	733 185 L	
Total volume of water added (simulation)	733 364 L	733 942 L
Control Effort (*CE*)	100.00 %	33.61 %

Table 3-2: Summary of simulation results for the time period from August 19th 0:00 to August 20th 23:00, 1995.

$$\%CE = \frac{CE_{\text{Filtered Data}}}{CE_{\text{Raw Data}}} \times 100\% \tag{3.4}$$

The considerable reduction in the control effort, when filtered feed rate data is used, implies that the amount of movement of the control valve will also be reduced by that amount over what is currently the case. This will result in a considerable reduction in the wear and tear of the tumbler water flow control valve. It must be pointed out here, that the reduction in control effort shown by this simulation represents a very conservative estimate of what can be achieved in the plant by filtering the feed rate signal. The reference data used here was collected at one minute intervals. Therefore, any control fluctuations with a period of less than two minutes are not visible in this data set, and therefore are not included in the calculation of $CE_{\text{Raw Data}}$. This results in a low estimate for this value. However, these more rapid fluctuations would also be removed by the low pass filter.

Figure 3-12 shows a plot of the tumbler water flow rates for the raw and filtered data case. For clarity, the scales for the two water rates have been shifted by 100 $\frac{L}{\text{sec}}$. The scale on the left is associated with the raw feed rate case (lower trace) and the scale on the right with the filtered data case (upper trace). The reduction in control effort is quite evident in this graph. The upper trace shows a significant reduction in high frequency variations while still matching the water rate closely with true shifts in feed rate.

3.5. Conclusion

Spectral analysis of process data offers a unique way of pinpointing periodic abnormalities in a set of data. In the case of the oil sand feed rate measurement, it was possible to identify the source of the spurious signal. Eliminating the spike with a period of four minutes in the frequency spectrum by applying a low pass filter to the

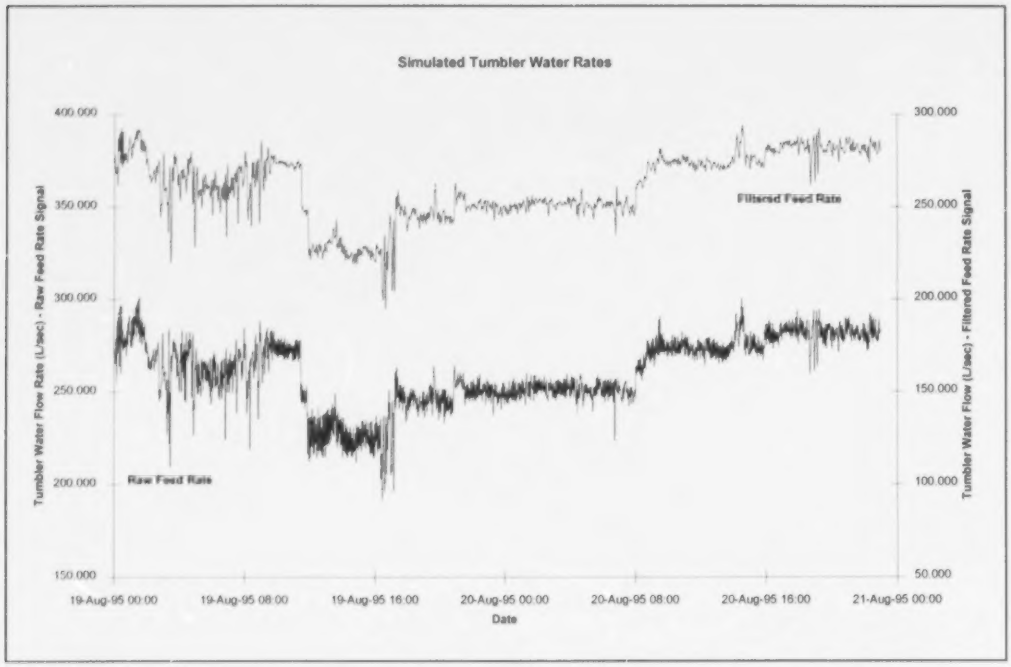

Figure 3-12: Comparison of tumbler water flow rates based on unfiltered and filtered feed rates.

signal resulted in a significant reduction in the control effort. This is expected to increase the service life of the water flow control valve substantially. It will result in a direct saving in maintenance cost without requiring any expenditures, since the ability to filter incoming signals is built into the digital control system. It has been shown through simulation, that the oil sand to water ratio is not affected by the feed rate data filter and that all 'true' fluctuations in feed rate are still matched by the control loop.

4. Time delay estimation for PSV model

Estimating the time delay between the inputs to a plant and its outputs is an important and necessary step not only in the development of a model, but is also essential when designing a controller for a process. Many modeling techniques can not accommodate time delays between the input and output data. It is therefore necessary to shift data in the input block with respect to each other and with respect to the output data, in order to arrive at a data set that has all time delays removed before trying to model a plant.

The data set for modeling the primary separation vessel (PSV) in the Syncrude extraction plant consists of 11 inputs and 4 outputs as shown in the block diagram in Figure 4-1. A process schematic outlining the physical location of the measurements is given in Figure 4-2. The difficulty in finding a suitable data set for this

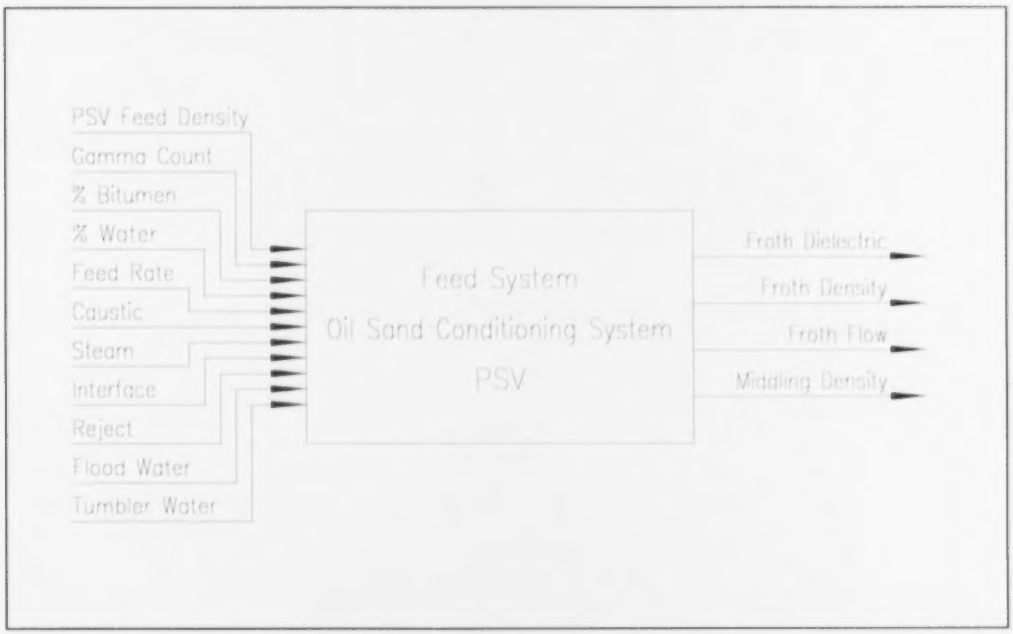

Figure 4-1: Block diagram of primary extraction circuit.

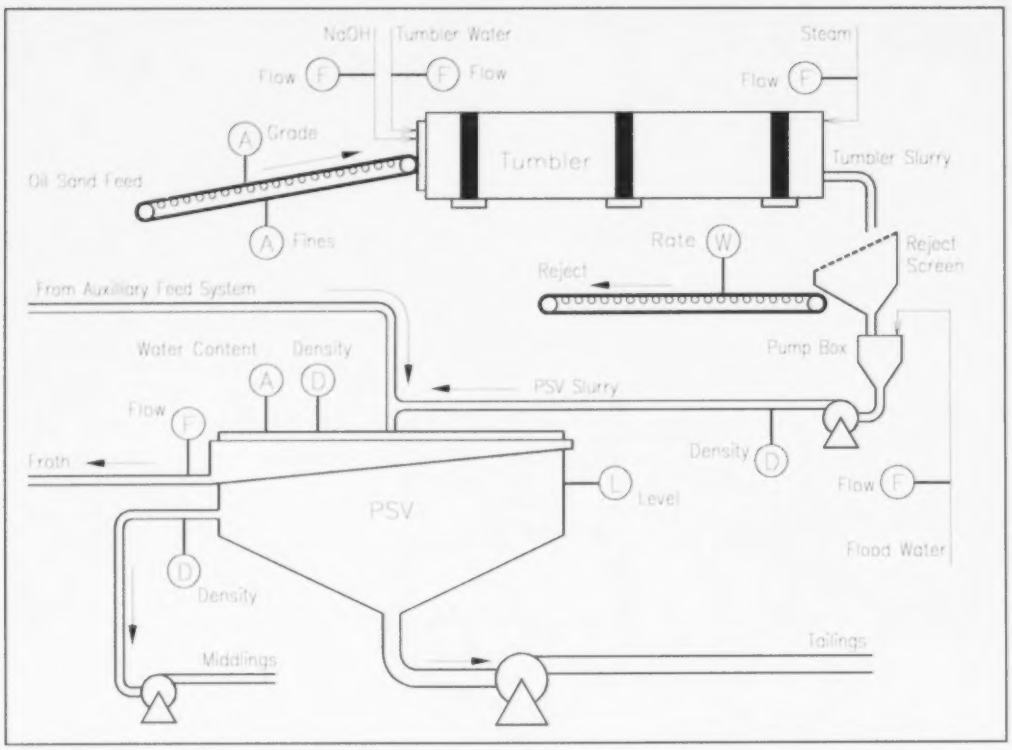

Figure 4-2: Schematic for the primary extraction circuit showing the physical location of the sensors.

investigation is that the PSV has two sources for it's feed slurry. One is the original feed system shown in Figure 4-2 and the other is the auxiliary feed system (see Figure 1-1 for a complete schematic). The auxiliary feed system can account for up to 25% of the total material entering the PSV. Unfortunately, it is very poorly instrumented and neither composition nor flow rate data is available for this stream. The implication of this is, that if a data set were used from a time when this feed system was operating, any model developed with this data set would be distorted by this very significant unmeasured 'disturbance'. When scanning the plant data base for a data set that could be used for deriving a model for the PSV, five days were found when the auxiliary feed system was shut down. These were four consecutive days from September 4th to September 7th, 1995, and 24 hours on September 9th, 1995. Each 24

hour period consists of 1440 data values for each variable. Tables 4-1 and 4-2 give a brief description of the input and output data respectively. Due to the inaccuracy of the two oil sand feed rate measurements, as explained in Chapter 2, the calculated feed rate, based on the material balance equation presented in the same chapter, was used as the feed rate input.

The challenge in determining time delays for a multiple-input system is that

Measurement	Description
% Bitumen	Bitumen content of oil sand feed as measured by the OSM.
% Water	Water content of oil sand feed as measured by the OSM.
Gamma Count	Measurement of naturally occurring γ radiation (Potassium 40) from oil sand – related to clay content of the oil sand feed.
Oil Sand Feed Rate	Calculated, based on PSV feed density.
PSV Feed Density	Density of slurry feed to the PSV.
Tumbler Water	Water flow rate into the tumbler.
Flood Water	Water flow rate into the pump box.
Caustic Flow	Flow rate of NaOH solution into the tumbler.
Steam	Steam addition rate to the tumbler.
Interface	Interface level between froth and middling layers inside the PSV.
Reject Rate	Rate of reject material removed from the tumbler slurry

Table 4-1: Summary of process inputs.

Measurement	Description
Froth Dielectric	Proportional to the water content of the froth.
Middling density	Density of the water layer below the froth in the PSV.
Froth density	Density of the froth at the top of the PSV.
Froth flow	PSV product flow rate.

Table 4-2: Summary of process outputs.

the time delay estimate must be done on a one input to one output basis. In this case, all other inputs, which may or may not influence a given output must be ignored. This makes determining the delay for that particular input – output pair difficult, since the interaction between the two signals may be masked by the other inputs. Four methods were used here to determine the time delays: cross-correlation, impulse response coefficient estimation, transportation delay estimation and a hybrid method combining the first two with the transportation delay method.

4.1. Use of cross-correlation for time delay estimate

The first method evaluated here was cross-correlation.[1, 2] It compares how well the input and output signals match when they are shifted with respect to each other one sample interval at a time. The number of shifts (lags) with the largest cross-correlation value represents the time delay in number of sample intervals between the two signals. To determine the time delays for the PSV data, a cross-correlation analysis for each input with each output (for a total of 44) was carried out for each of the 5 days.

To get an idea for what kind of results could be expected from this method, a simple simulation of a second order system with delay, including coloured noise, was carried out. The results of this simulation are plotted in Figure 4-3. The transfer function used is as follows:

$$y = \frac{0.75q^{-5} + 0.5q^{-6}}{1 - 0.25q^{-1} + 0.5q^{-2}} u + noise \tag{4.1}$$

It can be quite clearly seen that the number of lags, and hence the time delay of the system, where the highest cross-correlation value occurs is at 5 sample

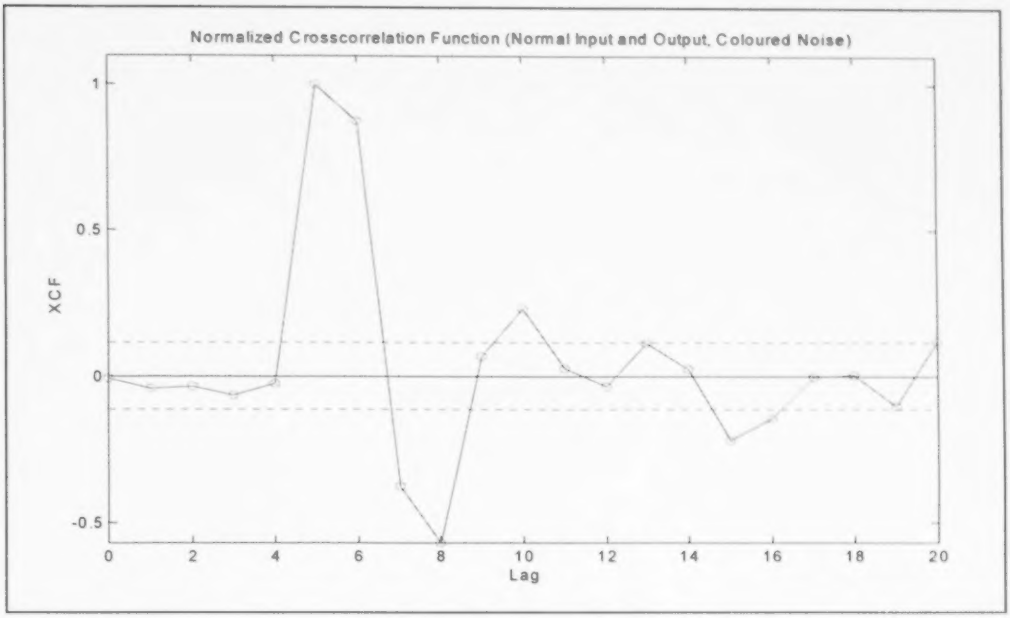

Figure 4-3: Cross-correlation results for simulation of second order system with delay and coloured noise.

intervals. The cross-correlation function used here is normalized, therefore for N number of data points in each data vector, the highest value over the entire range from $-N$ to $+N$ number of lags will be 1.0. Even though the cross-correlation is carried out over this range, only the region of interest, from 0 to 20 lags, is shown in Figure 4-3.

Applying the identical analysis to the plant data yielded quite different results. In the past, it had been observed that the clay content of the oil sand (γ count) had an effect on froth quality (froth dielectric). Of the five days examined here, only 3 gave any indication of a correlation between the two signals. Figures 4-4 and 4-5 are an example of this. They show the cross-correlation results for September 4th and September 5th respectively. The very broad peak in these plots are due to the influences of some of the other inputs on the froth dielectric reading. For September 4th, the curve reaches 1.0 at lag 22, giving a 22 minute time delay. For September 5th,

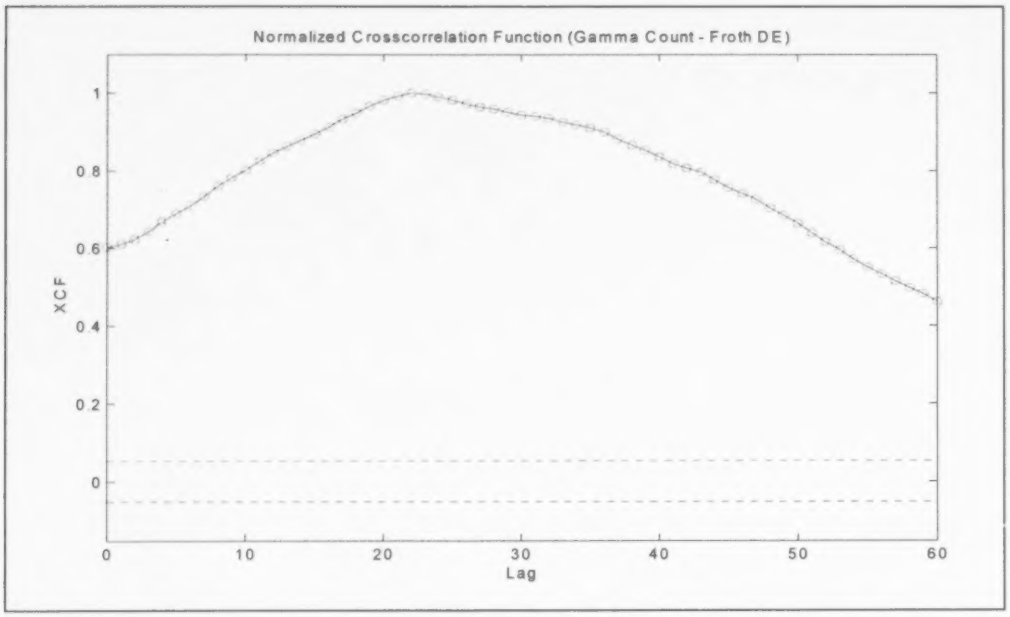

Figure 4-4: Cross-correlation result for γ count and froth dielectric for September 4th, 1995

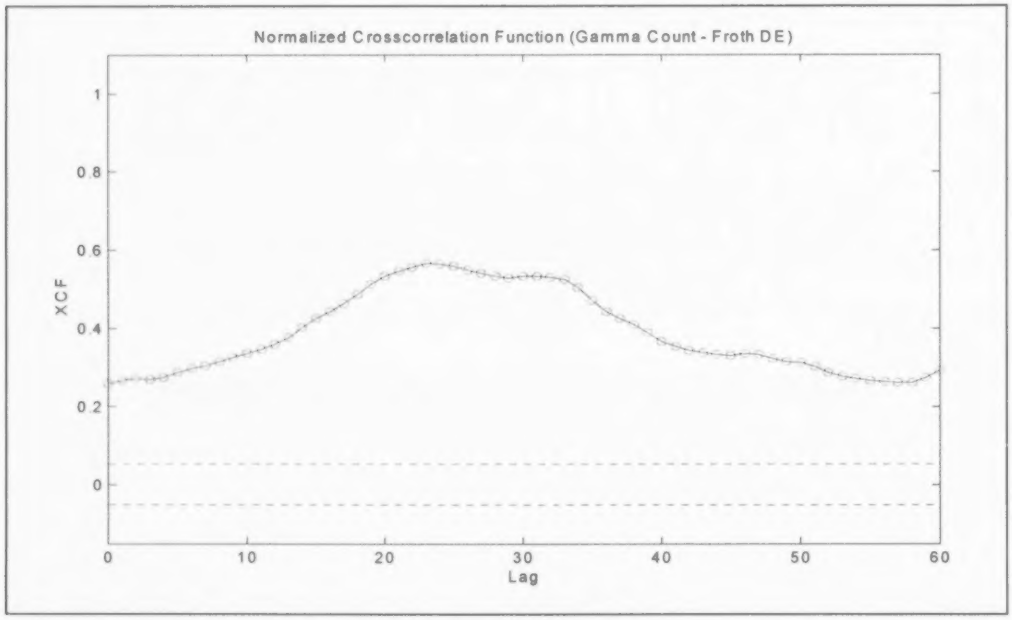

Figure 4-5: Cross-correlation result for γ count and froth dielectric for September 5th, 1995

however, the maximum value within the region of interest is just under 0.6, rather than the expected 1.0, indicating that the highest correlation between the two data vectors is outside the physically feasible region for this process. For the other two days that gave estimates for the time delay, the results were 29 minutes (September 7th) and 25 minutes (September 9th). Even though these differences appear to be quite large, these results are in better agreement than many others in this data set. Very few input – output combination gave consistent results. Table 4-3 summarizes the results of all cross-correlation time delay estimates. Blank cells within this table indicate that no time delay estimate could be obtained.

4.2. Use of impulse response coefficient estimation for time delay determination

Another method investigated was the use of impulse response coefficient determination to obtain the time delays.[3] When these coefficients are plotted, the lag of the first coefficient that falls outside the confidence interval around zero represents the delay of the system.

A simulation using the transfer function described in the previous section was performed for this method as well using the transfer function given in Equation 4.1. A plot of the impulse response coefficients resulting from this simulation is given in Figure 4-6. Again, the 5 minute delay for this system is unmistakable.

As with the cross-correlation method, the plots for the process data were quite different from the simulation. First of all, using the raw process data did not give any sensible results. Each input – output pair in the data set gave essentially the same result (Figure 4-7): the largest impulse response was at zero lag, indicating no time

Input	Output	September 4th	5th	6th	7th	9th
% Bitumen	Froth Density	15				15
% Water in Oil Sand	Froth Density	75				
γ Count	Froth Density					
Caustic Rate	Froth Density	15				
Feed Density	Froth Density	38				
Feed Rate (calc.)	Froth Density	22				32
Flood Water Rate	Froth Density					
PSV Interface Position	Froth Density					
Reject Rate	Froth Density					41
Steam Rate	Froth Density					
Tumbler Water Rate	Froth Density				7	
% Bitumen	Froth Dielectric	33				26
% Water in Oil Sand	Froth Dielectric					
γ Count	Froth Dielectric	22			29	25
Caustic Rate	Froth Dielectric	32			17	26
Feed Density	Froth Dielectric					
Feed Rate (calc.)	Froth Dielectric	51				
Flood Water Rate	Froth Dielectric					
PSV Interface Position	Froth Dielectric					
Reject Rate	Froth Dielectric			55		
Steam Rate	Froth Dielectric					
Tumbler Water Rate	Froth Dielectric					
% Bitumen	Froth Flow					16
% Water in Oil Sand	Froth Flow			8		
γ Count	Froth Flow	72			70	
Caustic Rate	Froth Flow	84			65	
Feed Density	Froth Flow		9		38	30
Feed Rate (calc.)	Froth Flow			46		33
Flood Water Rate	Froth Flow					
PSV Interface Position	Froth Flow		37			
Reject Rate	Froth Flow	44	60		39	58
Steam Rate	Froth Flow					
Tumbler Water Rate	Froth Flow				37	27
% Bitumen	Middling Density			14		
% Water in Oil Sand	Middling Density					
γ Count	Middling Density		20		25	16
Caustic Rate	Middling Density				7	9
Feed Density	Middling Density	5	6	36	4	
Feed Rate (calc.)	Middling Density	9	15		10	12
Flood Water Rate	Middling Density					
PSV Interface Position	Middling Density		34			
Reject Rate	Middling Density	36	14		5	7
Steam Rate	Middling Density					
Tumbler Water Rate	Middling Density				5	9

Table 4-3: Time delay estimation results (minutes), based on cross-correlation.

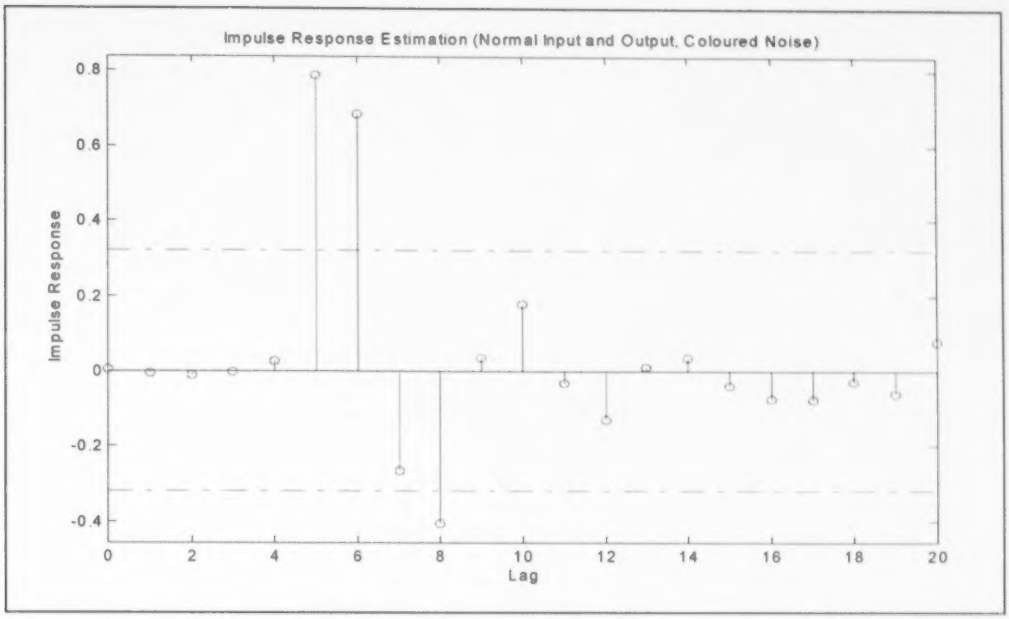

Figure 4-6: Impulse response coefficients for a simulated second order system with delay and coloured noise.

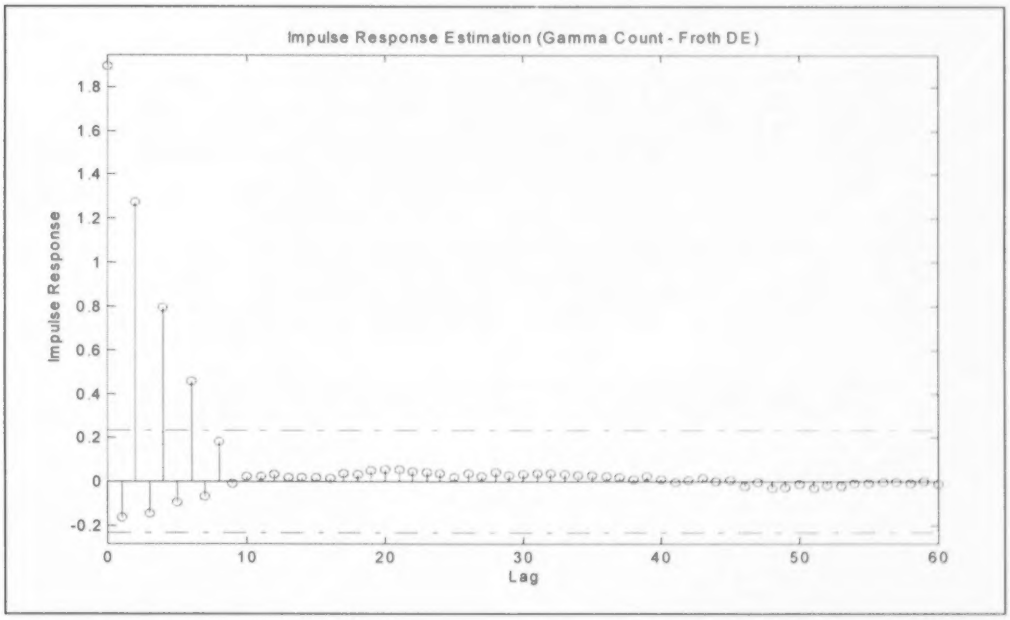

Figure 4-7: Impulse response coefficients for γ count and froth dielectric for September 4th, 1995, using raw process data.

delay whatsoever and subsequent coefficients alternate between positive and negative values and their magnitude was decaying exponentially.

In an attempt to improve the impulse response coefficient determination, the data set was normalized using the following equation:

$$x_i^* = \frac{x_i - \bar{x}}{\sigma_x} \tag{4.2}$$

where: x_i = raw data value
x_i^* = scaled data value
$\bar{x}$ = mean value of the data set
σ_x = variance of the data set

Repeating the impulse response determination using normalized data yielded results that were closer to what was expected. Unfortunately these were no more consistent than the cross-correlation results. Figure 4-8 shows the impulse response

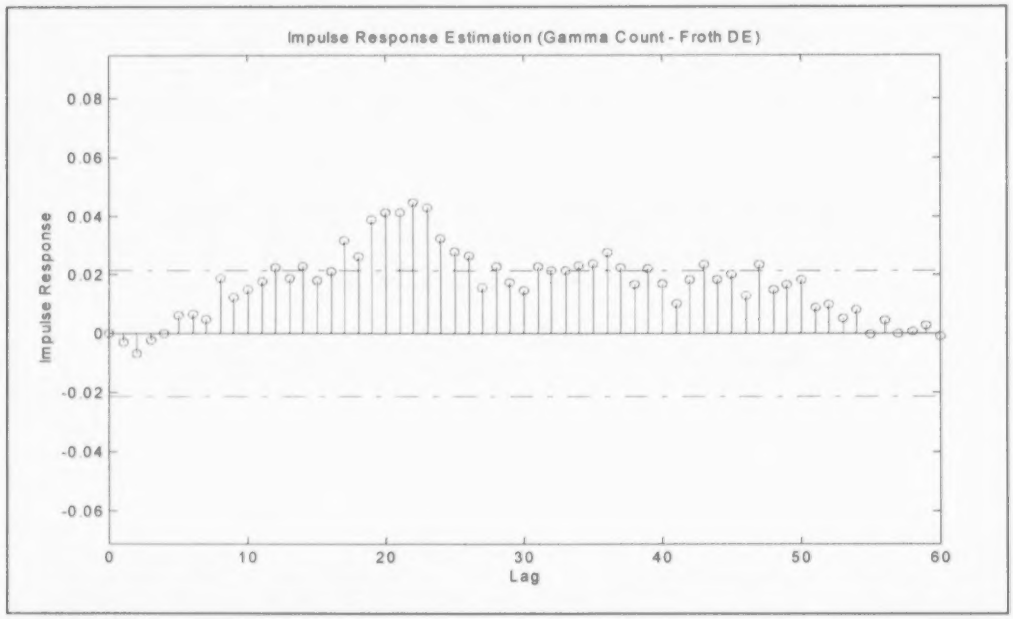

Figure 4-8: Impulse response coefficients for γ count and froth dielectric for September 4th, 1995, using normalized process data.

coefficients for the γ count and froth dielectric input – output pair for September 4th. The first significant (non zero) impulse response coefficient occurs at a lag of 17. For the September 5th data (Figure 4-9) no significant impulse response coefficients were found. For the other days, only September 7th (16 minutes) and September 9th (14 minutes) gave significant impulse responses. The fact that the days that did not yield any results for this input – output pair are the same as for the cross-correlation method is purely coincidental. Many incidences were found were one method gave a time delay estimate and the other did not. Table 4-4 summarizes the results of all impulse response based time delay estimates.

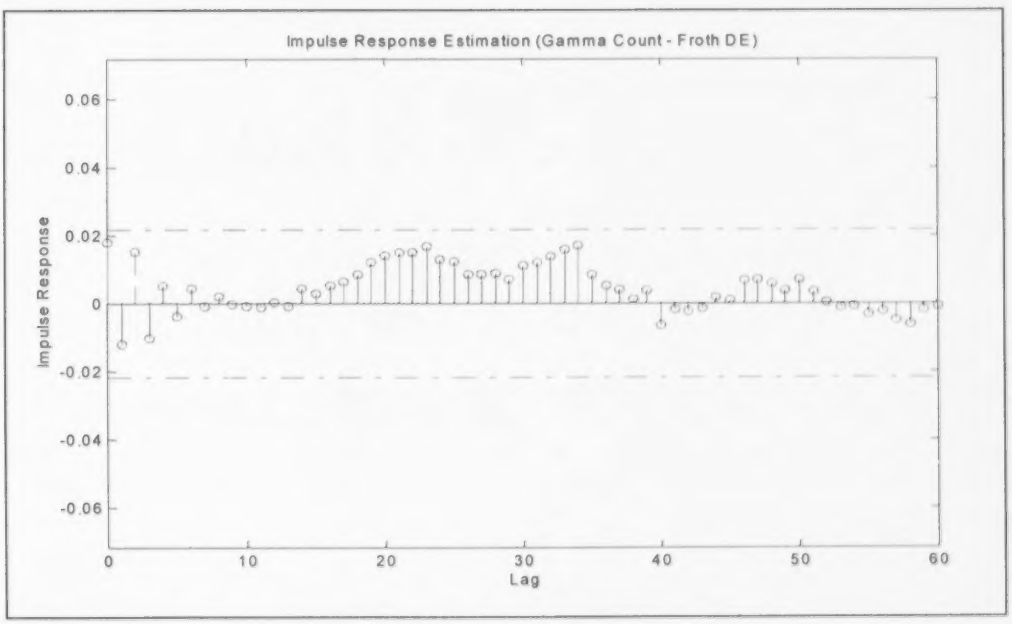

Figure 4-9: Impulse response coefficients for γ count and froth dielectric for September 5th, 1995, using normalized process data.

Input	Output	September 4th	5th	6th	7th	9th
% Bitumen	Froth Density		49			15
% Water in Oil Sand	Froth Density					57
γ Count	Froth Density		57	5	35	10
Caustic Rate	Froth Density					
Feed Density	Froth Density				14	12
Feed Rate (calc.)	Froth Density					17
Flood Water Rate	Froth Density		31	11	44	
PSV Interface Position	Froth Density		13			
Reject Rate	Froth Density				24	
Steam Rate	Froth Density		12	29		4
Tumbler Water Rate	Froth Density			11		
% Bitumen	Froth Dielectric	42		8	10	9
% Water in Oil Sand	Froth Dielectric			17		7
γ Count	Froth Dielectric	17		20	16	14
Caustic Rate	Froth Dielectric					
Feed Density	Froth Dielectric	3				10
Feed Rate (calc.)	Froth Dielectric	46			8	5
Flood Water Rate	Froth Dielectric					
PSV Interface Position	Froth Dielectric				3	11
Reject Rate	Froth Dielectric			32	23	25
Steam Rate	Froth Dielectric					3
Tumbler Water Rate	Froth Dielectric				7	
% Bitumen	Froth Flow	11			11	14
% Water in Oil Sand	Froth Flow			5		
γ Count	Froth Flow	11			54	
Caustic Rate	Froth Flow					
Feed Density	Froth Flow	37			1	4
Feed Rate (calc.)	Froth Flow	38				6
Flood Water Rate	Froth Flow					
PSV Interface Position	Froth Flow					2
Reject Rate	Froth Flow	37			1	
Steam Rate	Froth Flow	7				
Tumbler Water Rate	Froth Flow				9	
% Bitumen	Middling Density	8				9
% Water in Oil Sand	Middling Density					5
γ Count	Middling Density		12		5	6
Caustic Rate	Middling Density					
Feed Density	Middling Density	2		1	1	5
Feed Rate (calc.)	Middling Density	7		4	7	6
Flood Water Rate	Middling Density		6			
PSV Interface Position	Middling Density		7			
Reject Rate	Middling Density				3	21
Steam Rate	Middling Density		26			
Tumbler Water Rate	Middling Density	26				

Table 4-4: Time delay estimation results (minutes), based on impulse response.

4.3. Transportation time based time delay estimation

In contrast to the first two methods described here, this one does not rely on actual plant process measurements. It offers the advantage of being able to determine the delay between a particular input and output independently. For pure transportation delays (conveyor belts, pipes etc.) it is easy to arrive at a good estimate. However for delays that occur within a process unit, such as delays within the PSV, this method may give rather unreliable estimates. The transportation delays for the different sections of the process are given below using the Oil Sand Monitor location as the reference point:

- The transit time for the oil sand from the place where the Oil Sand Monitor and γ Analyzer are located on the plant feed conveyor to the tumbler entrance (front end) is 2 minutes.
- The inputs used for calculating the feed rate were adjusted to give the rate at the location of the Oil Sand Monitor, hence no time delay.
- The residence time of the material inside the tumbler is 3 minutes.
- The steam discharge pipes inside the tumbler run approximately two thirds of the length of the tumbler from the discharge end towards it's front, giving an average difference from the time the steam flow rate is measured until the water from the condensed steam is discharged from the tumbler of 2 minutes.
- The tumbler discharge to pump box time difference is negligible.
- The tumbler discharge to the reject rate measurement time

difference is 1 minute.

- The delay from the pump box to the PSV feed distributor is about 1 minute.

The time delays for the inputs are summarized in Table 4-5. Delays given in the table are with respect to the feed distributor inside the PSV and indicate how many minutes ago a particular input quantity was measured.

Estimating the delays inside the PSV is somewhat more tenuous. The time required for aerated bitumen droplets to rise through the middling layer to the froth – middling interface is, at the very least, a function of middling density, middling viscosity, turbulence inside the vessel and the amount of air attached to the droplet. These parameters are not known and probably vary considerably. It is therefore impossible to determine a delay for this part of the process. Another time delay that can not be ascertained by this method is the one between the feed distributor and the

Input	Time Delay (min.)
% Bitumen in oil sand feed	6
% Water in oil sand feed	6
γ count for oil sand feed	6
Oil sand feed rate	6
Tumbler water rate	4
Caustic rate	4
Steam Rate	3
Flood water rate	1
PSV feed density	1
Reject Rate	2

Table 4-5: Summary of transportation delays up to the PSV feed distributor.

middling density meter on the middling withdrawal line. However, the large volume of middling inside the PSV, in the order of 1 450m^3, results in very slow process dynamics, compared to which the time delay is most likely insignificant. Determining the rise time for bitumen through the froth layer is somewhat easier. If one assumes that the froth layer behaves as a plug flow, the average froth flow rate and the average froth volume inside the PSV can be used to estimate the time it takes for a particular bitumen droplet to travel from the middling – froth interface to the top of the froth layer and exit the PSV. This results in the following expression for the delay:

$$t_d^{Froth} = \frac{1000\,\pi (r_{PSV})^2 h_{Froth}}{\frac{\overline{m}_{Froth}}{\overline{\rho}_{Froth}}} \tag{4.3}$$

where: r_{PSV} = radius of PSV (9.45m)

h_{Froth} = nominal height of froth layer (0.8m)

$\overline{m}_{Froth}$ = average froth mass flow rate $\left(\frac{kg}{sec}\right)$

$\overline{\rho}_{Froth}$ = average froth density $\left(\frac{kg}{L}\right)$

The data and results for the time delay estimates for the froth layer are presented in Table 4-6. Considering the inconsistent estimates obtained by the previous two methods, the results in Table 4-6 are remarkably close. A value of 18 minutes will be used for overall time delays for this part of the process.

Date	Average Froth Flow $\left(\frac{kg}{sec}\right)$	Average Froth Density $\left(\frac{kg}{L}\right)$	Time Delay Estimate (min:sec)
September 4th	172.5	0.767	16:38
September 5th	159.2	0.831	19:32
September 6th	171.1	0.844	18.27
September 7th	164.2	0.821	18.43
September 9th	156.6	0.788	18:49

Table 4-6: Data and results for the froth layer time delay estimation.

Inputs	Outputs			
	Froth Dielectric	Middling Dens.	Froth Density	Froth Flow
% Bitumen	24	6	24	24
% Water	24	6	24	24
γ Count	24	6	24	24
Feed Rate	24	6	24	24
PSV Feed Dens.	19	1	19	19
Tumbler Water	22	4	22	22
Flood Water	19	1	19	19
Caustic Flow	22	4	22	22
Steam Rate	21	3	21	21
Interface				0
Reject Rate	20	2	20	20

Table 4-7: Matrix of time delays, in minutes, based on transportation delays only.

Since the top surface of the froth is fixed by the rim of the PSV, any movement in the froth – middling interface is reflected immediately in the froth flow rate, giving a zero time delay between the two measurements.

The results obtained thus far can be combined into time delay estimates for the process. A summary of these estimates is given in Table 4-7. It must be noted however that these estimates do not include the delays incurred in the middling layer, since these are impossible to estimate by this method.

4.4. Hybrid method for time delay estimation

One way of working around the limitations of a simple transportation delay approach to delay time determination is to combine it with one of the process data based determinations presented in the first two sections of this chapter. This approach would use the transportation delay method for the process up to the PSV feed distributor, where this method yield very reliable results, and use cross-correlation and/or impulse response coefficients for the PSV delays.

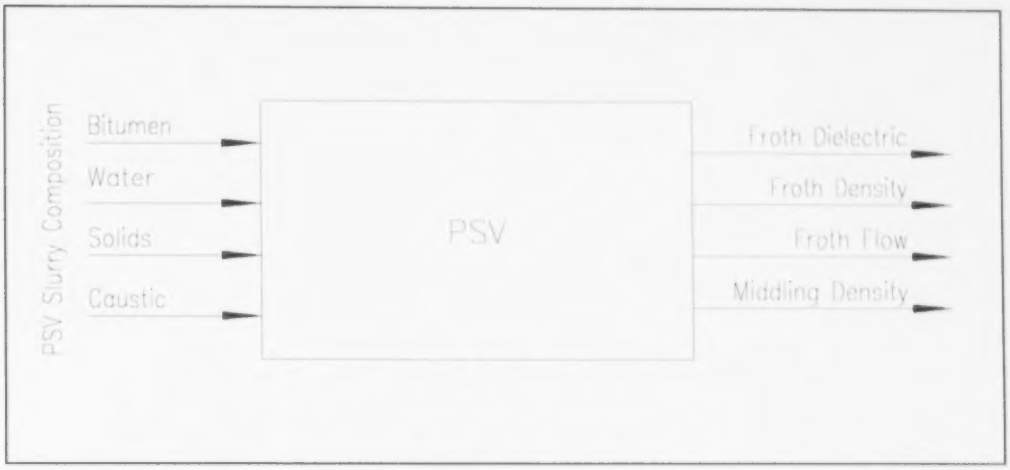

Figure 4-10: Block diagram of PSV input – output signals for PSV time delay estimation

The advantage of using the process data based methods only on the PSV is that instead of having to deal with eleven interactive inputs, the system is now reduced to four: namely the bitumen, water, solids and caustic which make up the PSV slurry. A block diagram for the PSV only is given in Figure 4-10. The slurry mass flow rates and composition can be obtained from the following material balance equations:

Mass flow rate of water in the PSV slurry:

$$m_W^{Slurry} = (W)(FR) + \rho_W^T(TW) + \alpha(SR) + \rho_W^F(FW) - (W)(m_R) \tag{4.4}$$

Mass flow rate of solids in the PSV slurry:

$$m_S^{Slurry} = (1 - W - B)(FR) - (1 - W - B)(m_R) \tag{4.5}$$

Mass flow rate of bitumen in the PSV slurry:

$$m_B^{Slurry} = (B)(FR) - (B)(m_R) \tag{4.6}$$

Mass flow rate of caustic in the PSV slurry:

$$m_C^{Slurry} = m_C - \frac{(W)(m_R)(m_C)}{(W)(FR) + \rho_W^T(TW) + \alpha(SR) + m_C} \tag{4.7}$$

Mass fraction of water in the PSV slurry:

$$x_W^{Slurry} = \frac{m_W^{Slurry}}{m_W^{Slurry} + m_S^{Slurry} + m_B^{Slurry} + m_C^{Slurry}} \tag{4.8}$$

Mass fraction of solids in the PSV slurry:

$$x_S^{Slurry} = \frac{m_S^{Slurry}}{m_W^{Slurry} + m_S^{Slurry} + m_B^{Slurry} + m_C^{Slurry}} \tag{4.9}$$

Mass fraction of bitumen in of the PSV slurry:

$$x_B^{Slurry} = \frac{m_B^{Slurry}}{m_W^{Slurry} + m_S^{Slurry} + m_B^{Slurry} + m_C^{Slurry}} \tag{4.10}$$

Mass fraction of caustic in of the PSV slurry:

$$x_C^{Slurry} = \frac{m_C^{Slurry}}{m_W^{Slurry} + m_S^{Slurry} + m_B^{Slurry} + m_C^{Slurry}} \tag{4.11}$$

The following notation was used in Equations 4.4 through 4.11:

m_C = mass flow rate of caustic to tumbler
m_R = mass flow rate of reject
m_C^{Slurry} = mass flow of caustic in slurry
m_B^{Slurry} = mass flow of bitumen in slurry
m_S^{Slurry} = mass flow of solids in slurry
m_W^{Slurry} = mass flow of water in slurry
x_C^{Slurry} = mass fraction of caustic in slurry
x_B^{Slurry} = mass fraction of bitumen in slurry
x_S^{Slurry} = mass fraction of solids in slurry
x_W^{Slurry} = mass fraction of water in slurry
ρ_W^F = density of flood water
ρ_W^T = density of tumbler water
α = steam condensation factor
B = bitumen fraction in oil sand
W = water fraction in oil sand
FR = oil sand feed rate
FW = flood water volume flow rate
SR = steam mass flow rate
TW = tumbler water volume flow rate

The results for the PSV delay estimates are given in Tables 4-8 and 4-9 for the cross-correlation and impulse response methods respectively. It is clear from these summaries, that the estimates, even though somewhat more consistent in some cases, are still not very reliable.

Input	Output	September 4th	5th	6th	7th	9th
Water mass flow	Froth Dielectric					
Bitumen mass flow	Froth Dielectric	32				16
Solids mass flow	Froth Dielectric					
Caustic mass flow	Froth Dielectric	28			14	23
% Water	Froth Dielectric					
% Bitumen	Froth Dielectric	29				15
% Solids	Froth Dielectric					
% Caustic	Froth Dielectric	28			19	22
Water mass flow	Middling Density					6
Bitumen mass flow	Middling Density		15	30		
Solids mass flow	Middling Density	4	7		5	
Caustic mass flow	Middling Density				3	6
% Water	Middling Density	5	9	34	4	
% Bitumen	Middling Density			7		
% Solids	Middling Density	5	7	35	4	13
% Caustic	Middling Density				3	6
Water mass flow	Froth Density				13	
Bitumen mass flow	Froth Density	11				1
Solids mass flow	Froth Density				6	27
Caustic mass flow	Froth Density	12		35		5
% Water	Froth Density	14				2
% Bitumen	Froth Density	11				2
% Solids	Froth Density	38			6	27
% Caustic	Froth Density	12		35		
Water mass flow	Froth Flow			36		25
Bitumen mass flow	Froth Flow			35		8
Solids mass flow	Froth Flow					28
Caustic mass flow	Froth Flow					
% Water	Froth Flow		7			30
% Bitumen	Froth Flow					8
% Solids	Froth Flow		6			30
% Caustic	Froth Flow			28		

Table 4-8: Summary of PSV time delay estimates (minutes) based on feed slurry composition using cross-correlation.

Input	Output	September 4th	5th	6th	7th	9th
Water mass flow	Froth Dielectric					4
Bitumen mass flow	Froth Dielectric	19				9
Solids mass flow	Froth Dielectric	19				14
Caustic mass flow	Froth Dielectric			15		
% Water	Froth Dielectric	19			3	10
% Bitumen	Froth Dielectric	19		13	5	3
% Solids	Froth Dielectric	30				10
% Caustic	Froth Dielectric	21			1	
Water mass flow	Middling Density		7	3		4
Bitumen mass flow	Middling Density	3		3	1	1
Solids mass flow	Middling Density	1	1	3	1	1
Caustic mass flow	Middling Density			30	1	
% Water	Middling Density	2	5	3	1	5
% Bitumen	Middling Density	3		3	3	5
% Solids	Middling Density	2	5	7	2	5
% Caustic	Middling Density				2	
Water mass flow	Froth Density		2			
Bitumen mass flow	Froth Density				24	1
Solids mass flow	Froth Density		2			1
Caustic mass flow	Froth Density					11
% Water	Froth Density			12		1
% Bitumen	Froth Density			12	24	1
% Solids	Froth Density					1
% Caustic	Froth Density					11
Water mass flow	Froth Flow					2
Bitumen mass flow	Froth Flow	6			6	1
Solids mass flow	Froth Flow	33			11	1
Caustic mass flow	Froth Flow					
% Water	Froth Flow	14				4
% Bitumen	Froth Flow	6			6	1
% Solids	Froth Flow	19			1	4
% Caustic	Froth Flow	20				

Table 4-9: Summary of PSV time delay estimates (minutes) based on feed slurry composition using impulse response coefficients.

4.5. Time delay variability

While some of the time delays for this process are constant, others exhibit significant fluctuations. In this section the different time delays will be examined for their variability individually.

The transportation delay for the oil sand on the conveyor belt is the same

regardless of how much material is being processed. It depends only on the speed of the conveyor belt which is constant. The same holds true for the reject measurement.

The residence time in the tumbler on the other hand is dependent not only on the oil sand feed rate, but also on the amount of tumbler water added. This process unit has, for all practical purposes, a fixed volume. Therefore, as the amount of material that is fed into the unit increases, the residence time decreases resulting in a shorted time delay. Fortunately, this variability is relatively small: no more than ± 1 minute over the normal feed rate range.

The reject screen and the pump box have almost no time delay whatsoever and thus do not contribute to the variability of the total delay in the system. Also, the delay in the slurry pipe between the pump box and the PSV, while having some dependency on the volume that passes through it, has a negligible contribution to the overall time delay fluctuations.

The part of the process that contributes the most to the changes in time delay is the PSV. Since it was not possible to determine delays for the middling layer, it is equally impossible to arrive at an estimate for it's changes. The plug flow model (Equation 4.3) discussed earlier, on the other hand lends itself well to a detailed analysis. The two factors that affect the delay in the froth layer are it's thickness and the froth flow rate. The nominal thickness of this layer is 0.8m, but it can easily vary by as much as ±0.2m. Using a constant froth flow rate equal to the average flow rate and average froth density recorded in the September 4th data set (172.48 $\frac{kg}{sec}$ and 0.7666 $\frac{kg}{L}$ respectively) the minimum, average and maximum time delay was calculated. The results are presented in Table 4-10. The second cause that was identified for changes in the dead time of the froth layer was froth flow rate.

Thickness of Froth Layer (m)	Froth Volume (L)	Time delay (minutes)
0.6	168,330	12.5
0.8	224,440	16.6
1.0	280,550	20.8

Table 4-10: Time delay variability due to changes in the froth layer thickness at a constant froth flow rate.

Assuming a constant froth thickness, Figure 4-11 shows the difference in the time delay based on the plug flow model for the September 4th data. The variations indicated here are also quite considerable ranging from 12.9 minute to 19.5 minutes. In reality the two variations discussed here separately combine giving a considerable uncertainty in any time delay estimate.

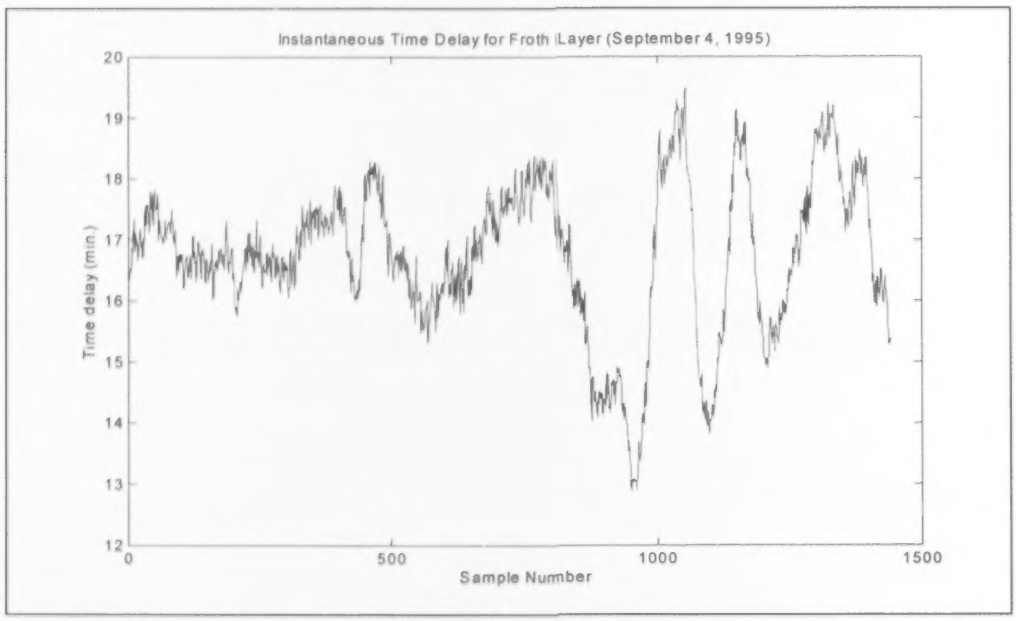

Figure 4-11: Time delay variability due to changes in froth flow rate at a constant froth layer thickness.

4.6. Conclusion

Finding a good method for time delay estimation is difficult for complex, highly interactive systems. The methods relying solely on process data yielded very inconsistent and inconclusive results, in part due to input – input interactions that can not be taken into account. The most significant cause for the poor results from the cross-correlation and impulse response estimation methods was the high variability in the actual time delays. The delay time for the froth layer is most likely the biggest single source of variability. In general, when results were obtained by both cross-correlation and impulse response estimation, the time delay from the latter method was the shorter of the two.

The transportation delay based method gave good delay estimates for the plant up to the PSV and also gave relatively close results for the 24 hour averages for the delay in the froth layer. However, no delays associated with the middling layer could be obtained. Quite possibly though, the delay within the middling layer is small relative to the time constant of any processes within it due to the very large volume of middling in the PSV (approximately 1 450m^3). and can be neglected. The analysis of the plug flow model for the froth layer did give some estimate of the short term variability in the time delays.

The hybrid method did not yielded much better results than the first two methods. Again most likely due to the extreme variability of the froth layer delays. Even though reducing the dimensionality of the system did only yield a small improvement in the consistency of the results for most input – output combinations, some of the values obtained, especially by the impulse response coefficient estimates, did show some similarity and supported the delays calculated for the PSV by the transportation delay method.

What this investigation clearly points out is that a 'black box' approach to time delay estimation in the case of a complex, highly interactive plant does not yield very reliable results. It is important, that any knowledge about the process is incorporated into the time delay estimation even though the possibility exists that bias is introduced into the estimate and that verifying the results through the use of process data may, at best, be difficult.

4.7. References

[1] Maybeck, P.S., "Stochastic Models, Estimation, and Control, Volume 1". Academic Press, Inc.

[2] Banerjee, P., "Robustness Issues in Long Range Predictive Control". Ph.D Thesis, University of Alberta (1996), pp 11-26.

[3] Box, G.E.P. and Jenkins, G.M., "Time Series Analysis Forecasting and Control". Holden-Day, Inc.

5. PLS Analysis of Extraction Data

The computer based control system in the Extraction Plant at Syncrude collects a vast amount of process data. Currently this information is only used by the operators for the immediate control of the process. All data is stored once per minute in a large data base. This data base represents a vast amount of information that, at present, is only sparsely utilized. The difficulty in analyzing this data, aside from the sheer number of entries (over 23,000 measurements per day for the part of the extraction circuit considered here alone, and in the neighborhood of 200,000 for all of extraction), lies in the complexity of the correlation between the measurements. In addition the reliability of many of the measurements is poor and often considerable noise is present in the data.

For a detailed description of the extraction process please refer to Chapter 1 of this thesis.

Attempts have been made in the past to develop models for the process based on mass balance equations, but many of the key measurements needed for them are either not available at all, or are only available on a very infrequent basis. Since the feed composition changes continually, frequent measurements are essential for defining a reliable model. Many of the parameters affecting the performance of the Primary Separation Vessel (PSV), which would allow for the calculation of product quality and quantity for example, can not be ascertained theoretically, and therefore are difficult to include in a theoretical model.

5.1. Objectives

The operating objectives for the extraction plant are quite simple: recover the

maximum amount of bitumen from the feed in the form of high quality (i.e. low water content) froth with a minimum amount of process inputs, such as caustic, hot water and steam. Currently there is very little automated help available to the operator to help him/her achieve this goal. They mainly rely on their past experience and feedback from the froth quality sensor to decide on operating parameters.

The objective of this study is to develop an empirical model of the process that will give the operator an indication of the quality of froth and the recovery of bitumen from the PSV for the current feed composition and operating conditions. The long time constants for the process and the very high throughput rates translate into a significant economic incentive for making this information available to the operators. For example, if the process conditions are not optimized for a particular feed grade, it will take over 30 minutes until the froth quality monitor will indicate poor froth quality and it will take an equal amount of time for the process to respond to the operator's corrective actions. During these 30 minutes, approximately 375,000 kg of poor quality froth will have been produced. A similar scenario holds true for bitumen recovery.

Due to the large number of variables and the high degree of correlation between them, many of the traditional statistical techniques, such as multiple linear regression analysis for example, either give very unreliable results or fail outright. Development of the Partial Least Squares (PLS) method was started by H. Wold in the mid 1960's and many refinements have been added subsequently by S Wold and coworkers.[1, 2, 3] Unlike Multiple Linear Regression (MLR), this method is capable of handling large, ill-conditioned, highly correlated and noisy data sets. In addition, it will identify the key variables that have the largest impact on the process output. The implication is that this method will take a high dimensional process and give the best

possible lower dimensional model. Identification of these key variables and how they affect the process means that the operator can control the process by monitoring only a few parameters closely instead of all parameters, a much more manageable task.

5.2. Sensitivities of the PLS method

Before applying PLS, or any other method for that matter, it is important to identify the limitations of the technique in the context of the available plant data. Most process identification methods, including PLS, are not able to cope with time delays (dead time) in the process. The input and output data must therefore be shifted to remove any dead time from the data set. Determination of the time delays for the extraction process has proven to be rather difficult (see Chapter 4) and the results obtained were not very reliable. Also, this method is essentially a steady-state identification technique. Since the oil sand processed in the extraction plant is a highly variable and non-homogeneous feed stock, the plant is, in reality, never at steady-state. A simple three input, one output model was used to examine these effects on the PLS results. In transfer function form, the model used in the following investigations is given by:

$$y = \begin{bmatrix} \dfrac{q^{-7}}{1-0.6q^{-1}} & \dfrac{q^{-5}}{1-0.75q^{-1}+0.5q^{-2}} & \dfrac{q^{-10}}{1-0.9q^{-1}} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \tag{5.1}$$

As can be seen from the above equation, this investigation was carried out on a noise free system.

5.2.1. Effect of non steady-state condition on PLS results

In order to examine this effect, one must compare the response of the process to the frequency content of the input signal. Figure 5-1 shows the step response for

the three individual transfer functions in the model. The top chart shows that the transfer function for u_1 has a very fast response time. The middle chart also exhibits a fast initial response, but requires somewhat more time to reach steady-state. The third transfer function, bottom chart, has the slowest response of the three. For the first simulation the input signals were designed in such a way that changes in the input did not occur any faster than twice the time required for their respective transfer functions to reach steady-state. Figure 5-2 gives the three input signals used in the first simulation.

The output for this system was simulated and the data was analyzed using PLS. Table 5-1 summarized the results of the analysis. Most of the variation is explained by the first latent variable, a much smaller amount by the second and virtually nothing by the third. A comparison of the PLS model output and the simulation is given in Figure 5.3. For the most part, the PLS predictions approximate

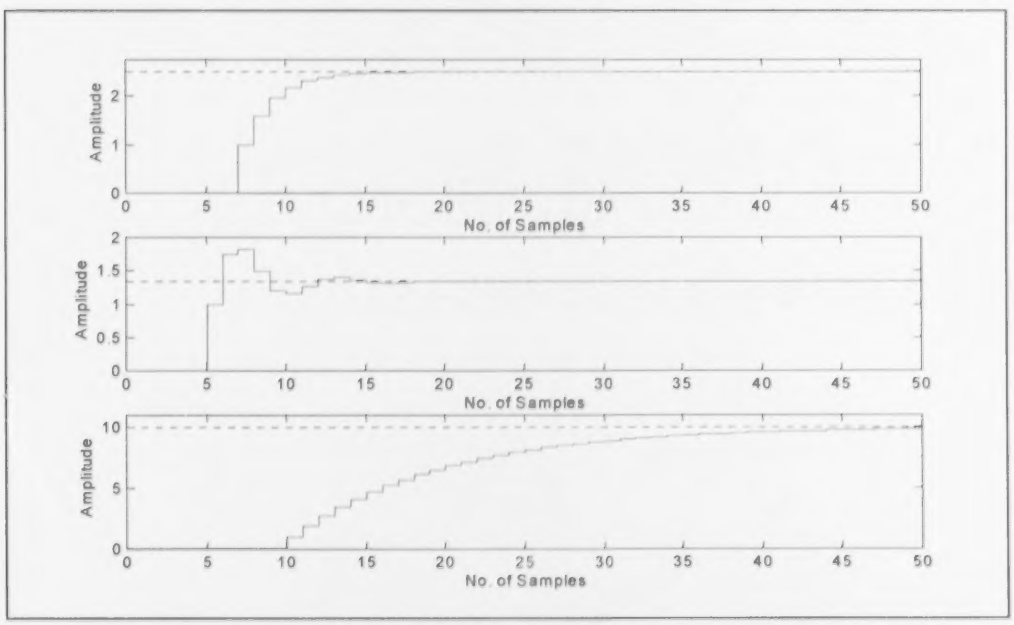

Figure 5-1: Step response for the three transfer functions used in the model.

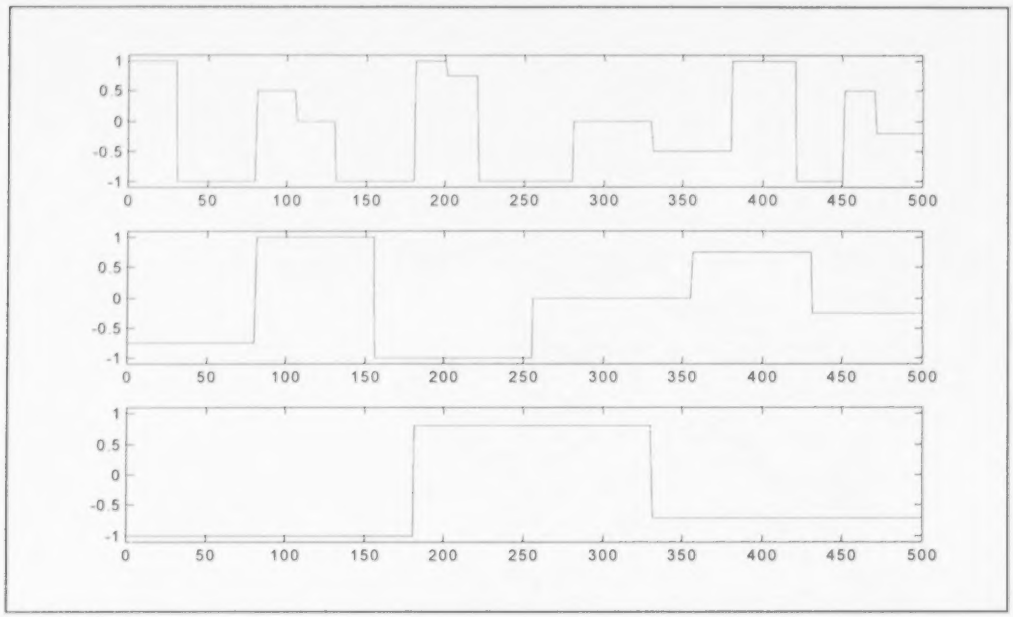

Figure 5-2: Plot of the three input signals used in the first simulation.

the output reasonably well. When comparing Figures 5-2 and 5-3, the three large changes in the level of the output can be attributed to u_3. The other variations can not be as clearly identified, but it appears that u_1 has a greater influence on y than u_2. Based on this simple comparison, the PLS results are not surprising. But it is also clear that, even in the absence of noise and with no unmeasured disturbances, this method can not capture all the variability in y, and therefore the total amount of output variance explained by all three latent variables is less than 100%.

Latent Variable Number	% Variance Captured by PLS Model
1	85.38
2	6.48
3	0.34
Total:	92.20

Table 5-1: PLS results for first simulation.

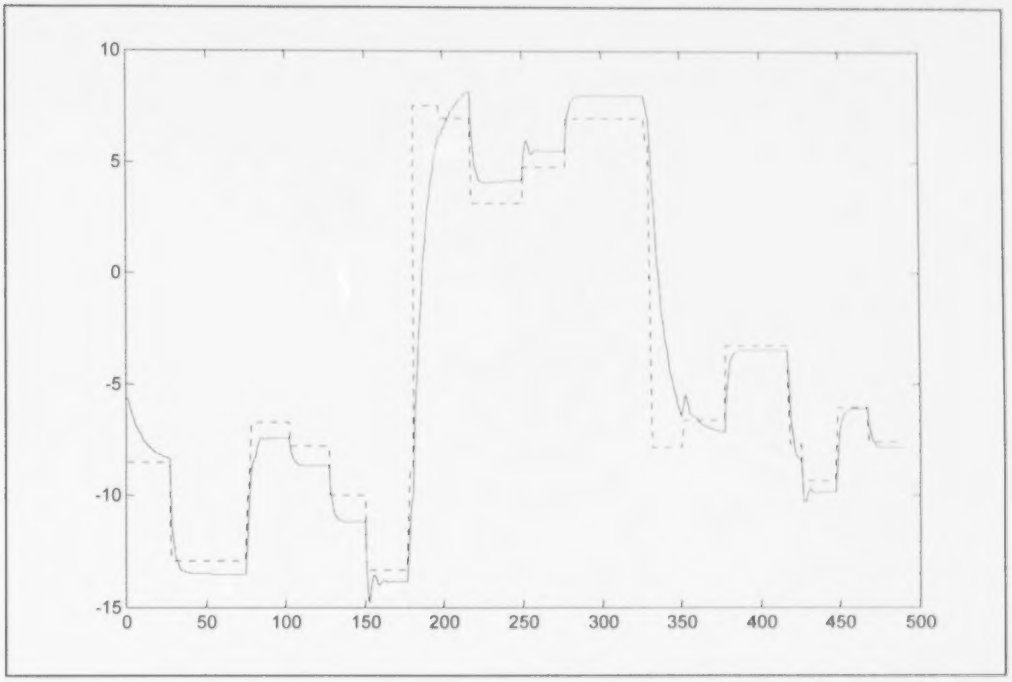

Figure 5-3: Comparison of output signal (—) for first simulation and PLS model predictions (---).

For the second simulation the number of transitions for all three inputs was increased to the point where in many cases steady-state could not be reached before the next step change in the input. Figure 5-4 shows three inputs, and Figure 5-5 gives the comparison of the output signal and the PLS model prediction. Similarities between the inputs and y are no longer clear. This is reflected in the PLS results (Table 5-2) for these signals. The large influence of u_3 on the output is still there, due to the much higher gain of it's transfer function, but is distorted by the fact that it was never able to attain steady-state. This lack of steady-state caused the PLS model to capture much less of the output variability. The method captures the 'level' changes quite well, but misses the transitions. Modifications to PLS have been proposed by J. Kresta to achieve dynamic identification.[4]

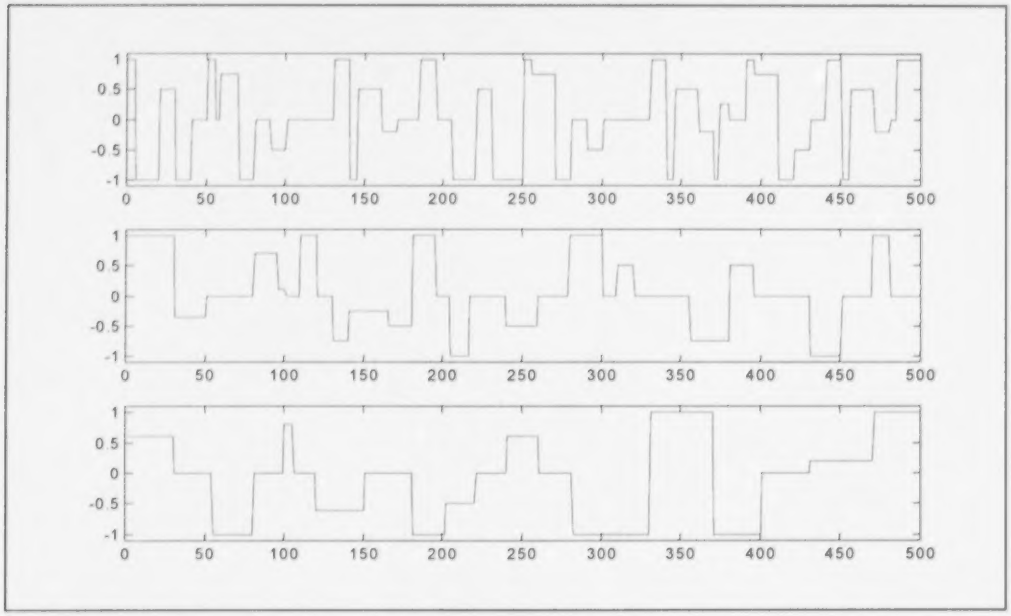

Figure 5-4: Plot of the three input signals used in the second simulation.

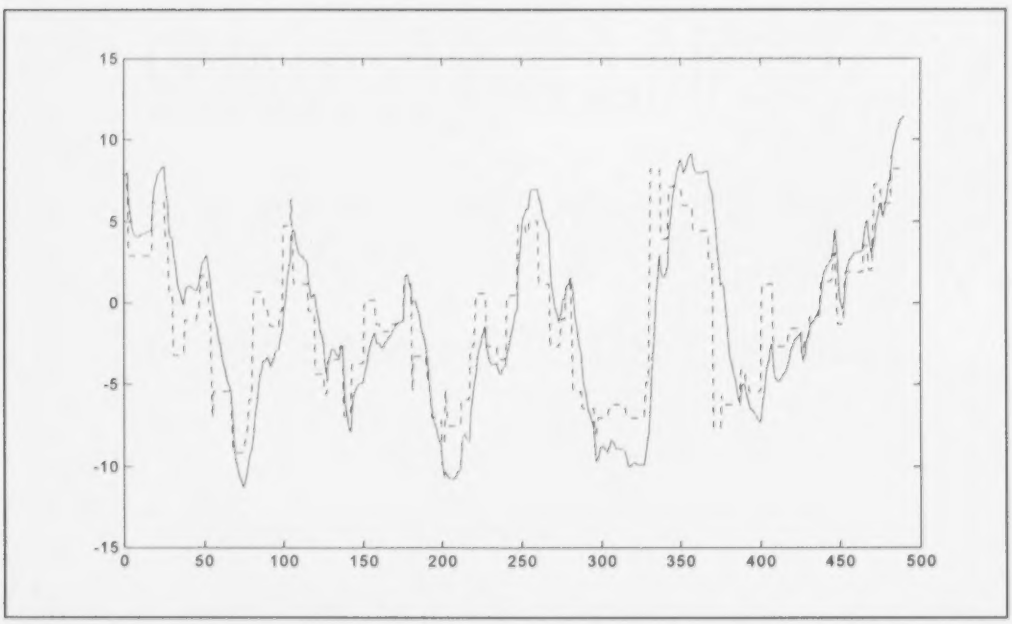

Figure 5-5: Comparison of output signal (—) for second simulation and PLS model predictions (---) using unfiltered data.

Latent Variable Number	% Variance Captured by PLS Model
1	68.96
2	1.10
3	0.00
Total:	70.06

Table 5-2: PLS results for second simulation using unfiltered data.

If the above hypothesis is true, then it would be expected that filtering the input and output signals with a filter that has a time constant that is just slightly longer than the slowest individual transfer function for the system, should restore quasi steady-state conditions and improve the predictive capabilities of the PLS model. For this example here, this task can be easily accomplished since the transfer functions are known. For actual process data, this is not the case and a method must be found for selecting a filter based only on the data. A common way of verifying how good a simulation result is, is to calculate the residuals (Equation 5.2) and check if they are uncorrelated (white).

$$resid(k) = y_{actual}(k) - y_{predicted}(k) \tag{5.2}$$

The residuals for the second simulation were determined. Using the autocorrelation function on them revealed a significant correlation pattern. The filter needed to convert the residuals to white noise was calculated by applying the autoregression function. The resulting unity gain filter function is given by:

$$\frac{0.0822}{1 - 0.9178q^{-1}} \tag{5.3}$$

The time constant for this filter is indeed just slightly longer than the one for the slowest transfer function. Using this filter on the above residuals did yield white noise. Figure 5-6 shows the autocorrelation for the unfiltered residuals on the top and for the filtered ones on the bottom.

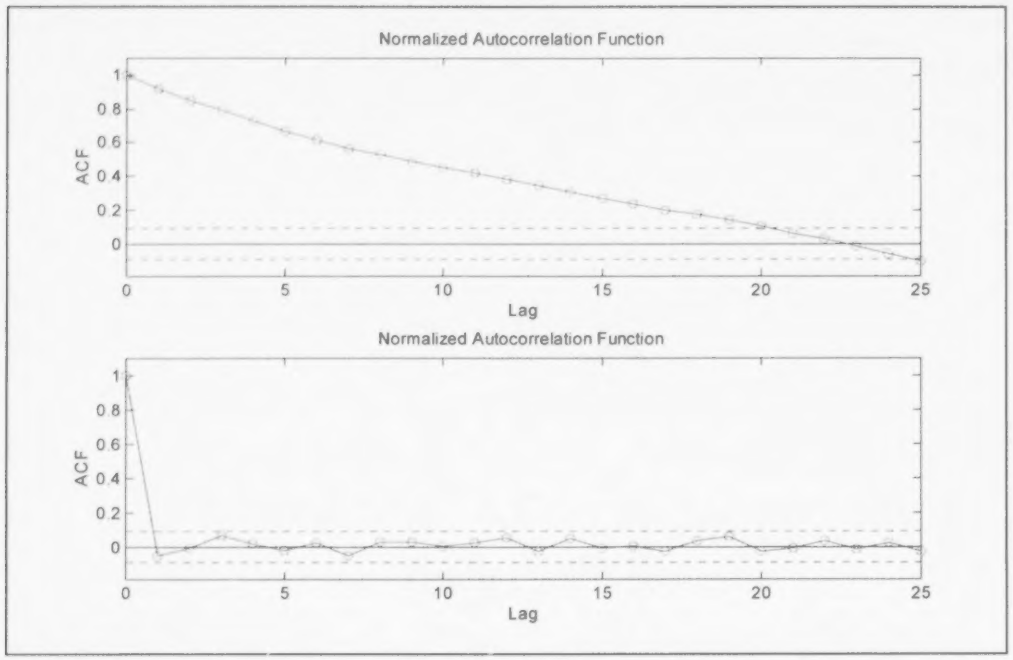

Figure 5-6: Comparison of unfiltered (top) and filtered (bottom) residuals from the second simulation.

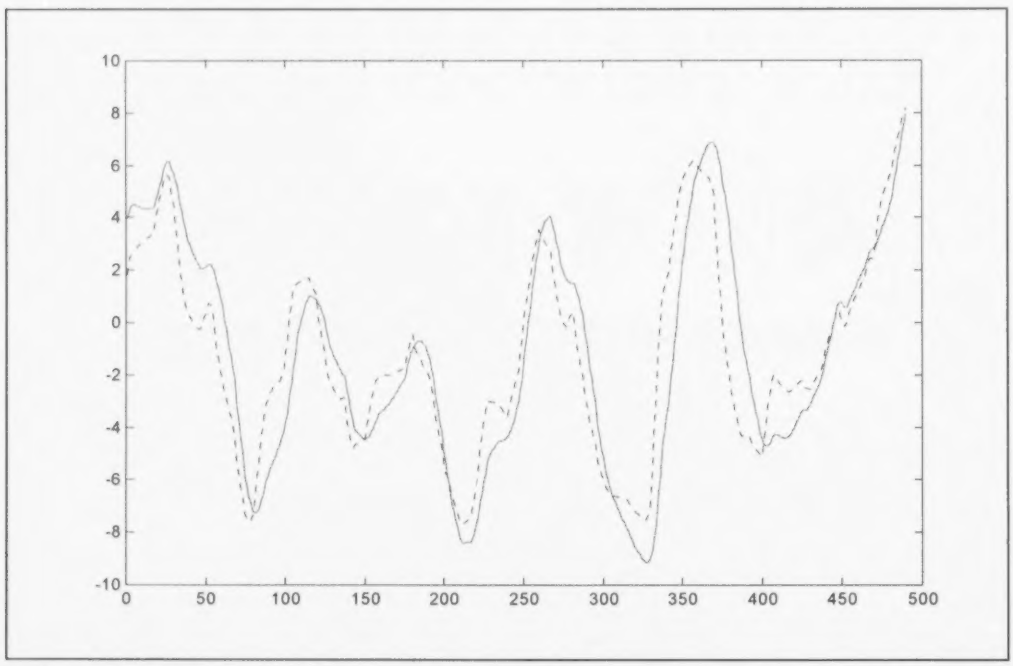

Figure 5-7: Comparison of output signal (—) for second simulation and PLS model predictions (---) using filtered data.

Latent Variable Number	% Variance Captured by PLS Model
1	76.17
2	2.18
3	0.01
Total:	78.36

Table 5-3: PLS results for second simulation using filtered data.

The next step was to filter the three inputs used in the second simulation and the corresponding output and repeat the PLS analysis using filtered data.. The results of this analysis are given in Table 5-3. While this did not give back the 92% total captured variance seen in the first simulation, it did increase the predictive capability of the PLS model. Figure 5-7 shows the results of the PLS model predictions using data that was filtered.

Even though more of the output variance is captured when the input and output data is filtered, this comes at the expense reduced time resolution. A much longer time constant filter was also tried $\left(\frac{0.02}{1-0.98q^{-1}}\right)$, which did increase the total variance explained to just over 86%, but the resolution of the model was further reduced. It is felt that the filter obtained by using the autoregression function on the residuals represents a good compromise between resolution and predictive capability for the extraction data.

5.2.2. Effect of time delay errors on PLS results

Obtaining an accurate estimate of the time delay for the process data proved to be quite difficult (Chapter 4). To evaluate the sensitivity of the PLS method to errors in the time delay, a simulation was carried out using the same model and the same inputs described in the previous section. From the transfer functions and their

Latent Variable Number	% Variance Captured by PLS Model		% Change Due to Error
	Correct Delays	1 Sample Interval Error	
1	68.96	60.93	11.6
2	1.10	0.95	13.6
3	0.00	0.00	0.0
Total:	70.06	61.88	

Table 5-4: PLS results for time delay error analysis (unfiltered data).

respective step responses, it can be seen that the correct time delays for the model are 7, 5 and 10 sample intervals for u_1, u_2 and u_3, respectively. The PLS analysis carried out in the second simulation above using unfiltered data and the correct time delays will be used for a reference. The output was the shifted by 1 sample interval changing the delays for the three inputs to 6, 4 and 9 and the analysis was repeated without filtering the data. The results are given in Table 5-4. When these results are compared, it can be seen that even the smallest possible error (one sample interval) has a significant effect on the result. Furthermore, the relative change in the first two latent variables is of the same magnitude.

The next step was to investigate the effect of filtering the input and output data with several different short to medium (approximately 3 to 10 sample periods) time constant filter to see if this reduces the sensitivity to time delay errors. The

Latent Variable Number	% Variance Captured by PLS Model		% Change Due to Error
	Correct Delays	1 Sample Interval Error	
1	70.17	62.83	10.4
2	1.17	1.03	12.0
3	0.00	0.00	0.0
Total:	71.34	63.86	

Table 5-5: PLS results for time delay error analysis (filtered data).

results for this are given in Table 5-5. Unlike in the previous section, where filtering gave improved model predictability, it did not reduce the effect of errors in the time delay estimate for the system appreciably.

5.3. Time delays for the extraction data

In light of the difficulties encountered in estimating the time delays for the extraction data sets described earlier, the system delays were re-examined using the time delay sensitivity of the PLS method. As a staring point for this analysis, the delays estimated from Chapter 4 were used. These were then varied, over what was felt to be a reasonable range for this process, and the total percent variance captured by the PLS model was calculated for each of the delay estimates. This was done for all five days of data. Tables 5-6, 5-7, 5-8 and 5-9 show the total percent variance explained by the PLS model for different time delays for froth density, froth flow rate, froth dielectric (quality) and middling density for the five days using the raw data. The **bolded** entries in the tables indicate the highest captured variance for each day. The delay time marked with an asterisk (*) is the one that was estimated in the previous chapter. Since a single time delay is required for building a process model and testing its validity and the new estimates still did show quite a bit of variability from one day to the next, the delays that gave the highest captured variance were simply averaged. These average delays, given in Table 5-10, were then used to build the model. One interesting point to note is that the variance captured by the model differs considerably more from one day to the next than it does as a result of the assumed time delays. One possible explanation is that this is due to relatively slow, but significant, unmeasured changes in either the feed or the process itself.

Time Delay (minutes)	Total % Variance Captured by PLS Model				
	Sept. 4th	Sept. 5th	Sept. 6th	Sept. 7th	Sept. 9th
14	**40.78**	**68.24**	70.01	45.59	38.42
16	40.56	66.79	70.02	**45.86**	37.61
20	38.21	66.53	**70.11**	44.81	40.18
24*	36.35	65.71	69.21	43.41	**41.63**
29	30.48	66.05	69.52	41.57	31.60
34	29.27	65.82	68.94	40.54	25.86

Table 5-6: Total percent variance captured by PLS model for froth density as a function of time delay.

Time Delay (minutes)	Total % Variance Captured by PLS Model				
	Sept. 4th	Sept. 5th	Sept. 6th	Sept. 7th	Sept. 9th
14	30.17	**49.96**	60.76	57.51	**38.80**
19	**30.73**	49.02	61.66	58.83	35.85
24*	29.68	46.96	**61.72**	59.56	35.76
29	26.58	43.97	61.05	62.14	34.17
34	25.02	43.09	61.36	64.59	29.58
39	24.17	43.92	61.21	**66.78**	24.39

Table 5-7: Total percent variance captured by PLS model for froth flow rate as a function of time delay.

Time Delay (minutes)	Total % Variance Captured by PLS Model				
	Sept. 4th	Sept. 5th	Sept. 6th	Sept. 7th	Sept. 9th
19	41.19	61.43	36.98	63.13	56.39
24*	50.73	62.01	36.91	**64.60**	**63.87**
29	56.57	**62.11**	**39.75**	64.06	51.51
34	**56.79**	59.75	39.35	61.30	47.19
39	52.17	56.82	38.60	57.15	41.83

Table 5-8: Total percent variance captured by PLS model for froth dielectric as a function of time delay.

Time Delay (minutes)	Total % Variance Captured by PLS Model				
	Sept. 4th	Sept. 5th	Sept. 6th	Sept. 7th	Sept. 9th
5	60.59	69.77	65.90	68.74	36.50
6*	61.11	70.66	66.44	69.91	37.39
7	61.41	71.53	66.96	71.22	38.28
9	**61.63**	73.16	68.04	**72.94**	40.83
11	60.84	74.66	68.80	72.48	**41.09**
14	59.04	76.13	**68.90**	70.07	40.19
16	57.72	**76.88**	68.06	68.21	39.31

Table 5-9: Total percent variance captured by PLS model for middling density as a function of time delay.

5.4. Determination of data filters for the extraction data

Using the time delays given in Table 5-10, a PLS analysis was carried out on the PSV data set for all 5 days. The residuals were calculated for each day's PLS model predictions, using the same data set for both model determination and verification. These residuals were then employed to determine the data filter as described in Section 5.2.1 above. The filter specified was a third-order, unity gain filter. Even though, when applied to the residuals, it did not result in perfectly white noise (one or two points fell outside the 95% confidence region around zero of the

Process Measurement	Time Delay (Minutes)
Froth density	18
Froth flow rate	22
Froth dielectric	28
Middling density	12

Table 5-10: Time delays used for the PLS model.

Sept.	Froth density	Froth flow rate	Froth dielectric	Middling density
4th	$\frac{0.16}{1-0.93q^{-1}+0.47q^{-2}-0.38q^{-3}}$	$\frac{0.03}{1-1.15q^{-1}+0.45q^{-2}-0.27q^{-3}}$	$\frac{0.06}{1-1.01q^{-1}+0.40q^{-2}-0.33q^{-3}}$	$\frac{0.07}{1-1.07q^{-1}+0.36q^{-2}-0.22q^{-3}}$
5th	$\frac{0.10}{1-0.99q^{-1}+0.51q^{-2}-0.42q^{-3}}$	$\frac{0.03}{1-0.72q^{-1}+0.02q^{-2}-0.27q^{-3}}$	$\frac{0.03}{1-1.24q^{-1}+0.62q^{-2}-0.35q^{-3}}$	$\frac{0.04}{1-0.99q^{-1}+0.31q^{-2}-0.28q^{-3}}$
6th	$\frac{0.12}{1-0.96q^{-1}+0.45q^{-2}-0.37q^{-3}}$	$\frac{0.03}{1-1.16q^{-1}+0.37q^{-2}-0.17q^{-3}}$	$\frac{0.06}{1-1.24q^{-1}+0.59q^{-2}-0.29q^{-3}}$	$\frac{0.05}{1-1.03q^{-1}+0.27q^{-2}-0.19q^{-3}}$
7th	$\frac{0.19}{1-0.97q^{-1}+0.50q^{-2}-0.34q^{-3}}$	$\frac{0.04}{1-1.13q^{-1}+0.47q^{-2}-0.29q^{-3}}$	$\frac{0.05}{1-1.39q^{-1}+0.70q^{-2}-0.27q^{-3}}$	$\frac{0.08}{1-1.17q^{-1}+0.30q^{-2}-0.05q^{-3}}$
9th	$\frac{0.13}{1-1.01q^{-1}+0.50q^{-2}-0.37q^{-3}}$	$\frac{0.03}{1-1.04q^{-1}+0.38q^{-2}-0.31q^{-3}}$	$\frac{0.05}{1-1.64q^{-1}+1.01q^{-2}-0.33q^{-3}}$	$\frac{0.04}{1-1.18q^{-1}+0.64q^{-2}-0.42q^{-3}}$

Table 5-11: Summary of data filters for the PSV data based on autoregression of the residuals.

autocorrelation function), it is a good compromise between complexity and effectiveness. The resulting filters for each day for the four outputs are given in Table 5-11.

Simply looking at the filter equations in Table 5-8 does not give a clear picture of how much they differ from one day to the next for a given output or between outputs. It is much easier to see the differences between the filters when their step responses are plotted. Figures 5-8, 5-9, 5-10 and 5-11 are the step responses for all five days for froth density, froth flow rate, froth dielectric and middling density, respectively. While there are day to day differences between the optimal filters these are not as much as the differences between some of the outputs. It is clear from these plots that any one of the filters can be used to filter the data before building a model for a given output, but that different filters should be used for the different outputs.

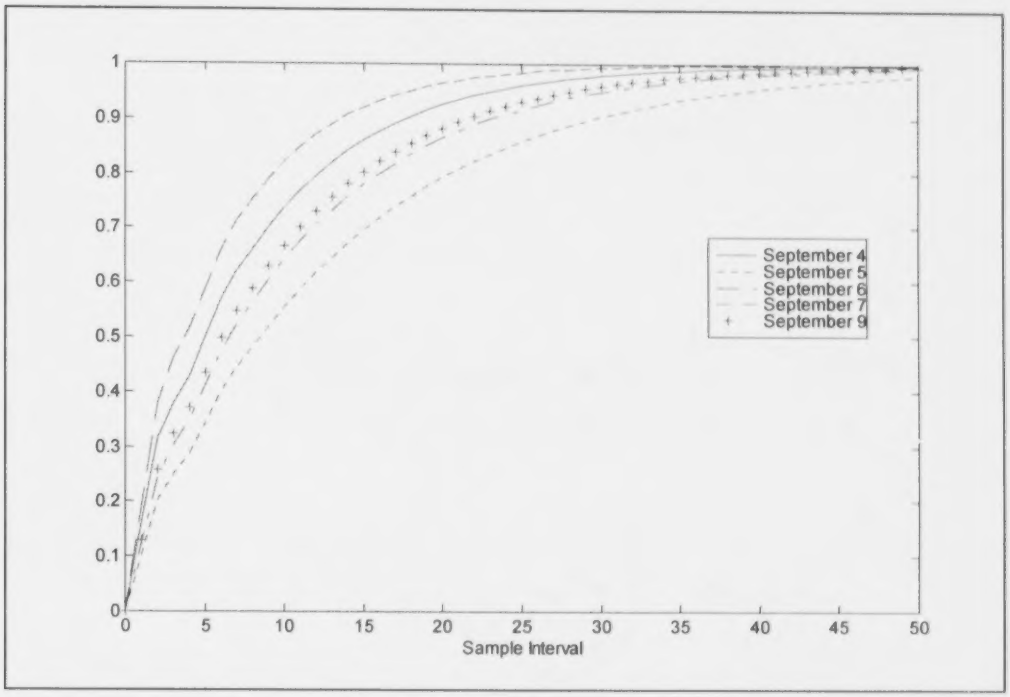

Figure 5-8: Step responses of the data filters for the froth density.

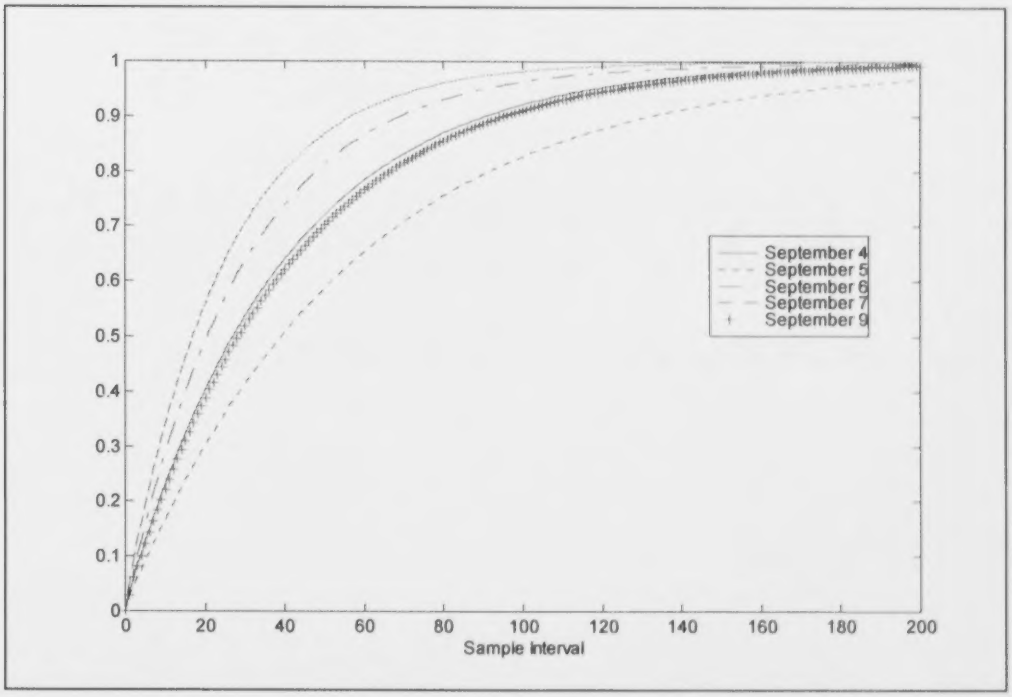

Figure 5-9: Step responses of the data filters for the froth flow rate.

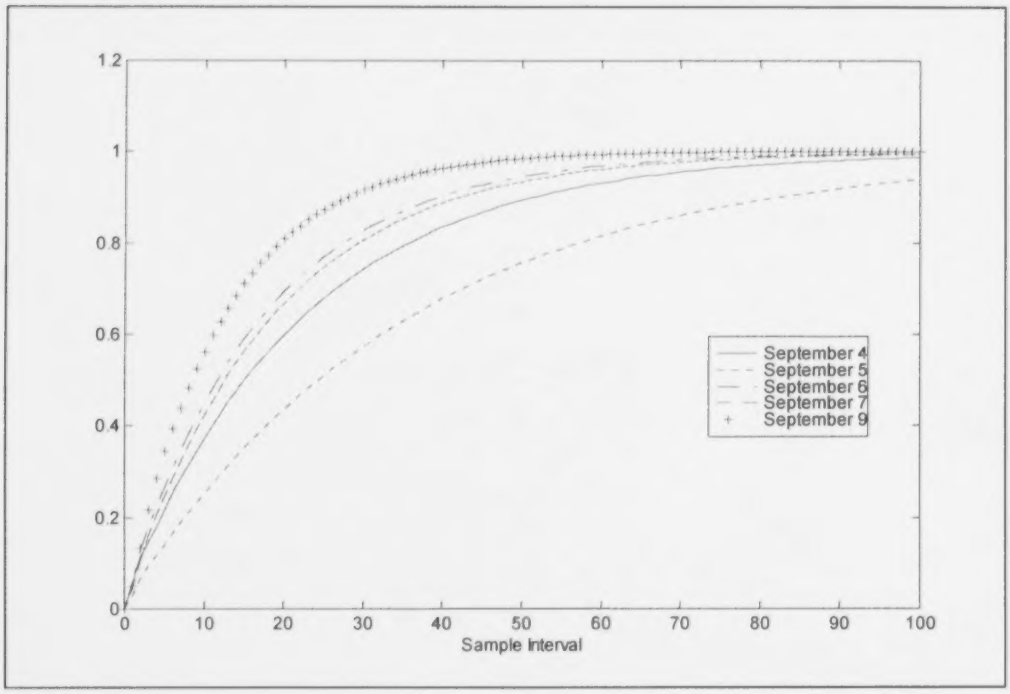

Figure 5-10: Step responses of the data filters for the froth dielectric.

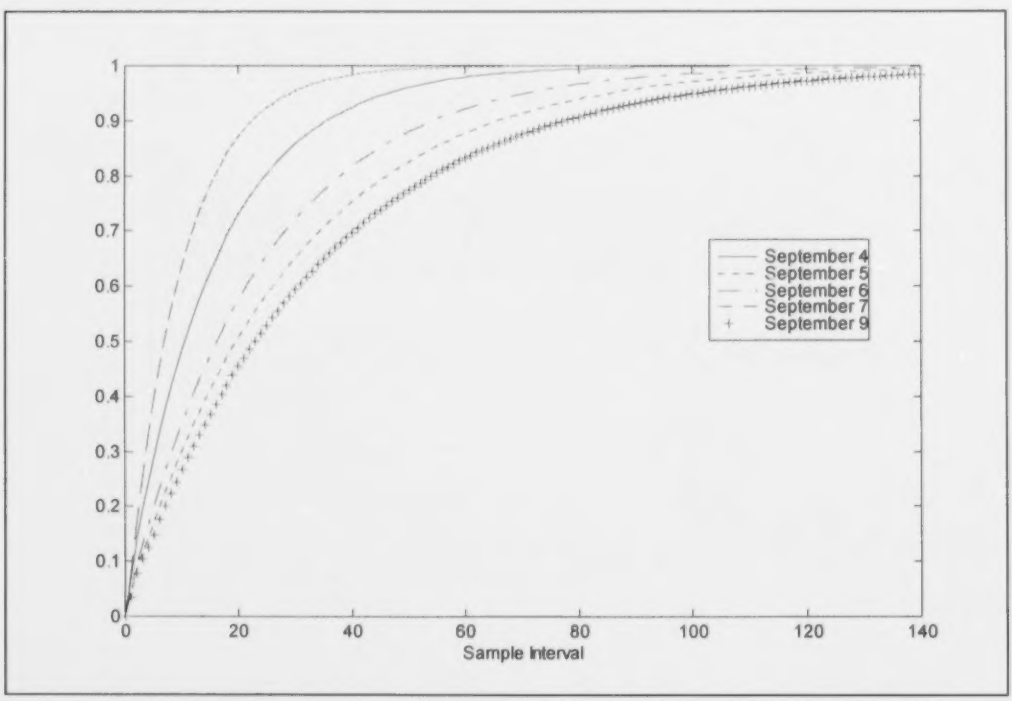

Figure 5-11: Step responses of the data filters for the middling density.

5.5. PLS model for the extraction data

The PLS method is capable of handling both multiple input / single output (MISO) and multiple input / multiple output (MIMO) systems. Any MIMO system can be divided into as many MISO systems as there are outputs. Using the MISO approach gives more freedom in tailoring the data filter for each individual MISO system, In addition, a different number of latent variables can be chosen for each. This allows for optimizing the analysis for each output. The MIMO approach on the other hand requires the use of a single data filter and one set of latent variables. However, it takes into account correlation between the different outputs, which can have a significant affect on the model. Both approaches will be used for the extraction data set.

From the investigations carried out thus far, it appears that the data set for September 5th is the most consistent. This is based on the total variance captured by the PLS model determined when testing the data sets for the time delay in Section 5.3. The PLS models will be developed using this data set.

5.5.1. Four separate PLS models (MISO approach)

Both the input and output data was filtered prior to model identification with the filters developed in the previous section for this particular data set, specifically:

Filter used for the froth density model:

$$\frac{0.1021}{1.000 - 0.9899q^{-1} + 0.5135q^{-2} - 0.4215q^{-3}} \tag{5.4}$$

Filter used for the froth flow rate model:

$$\frac{0.0267}{1.000 - 0.7237q^{-1} + 0.0195q^{-2} - 0.2690q^{-3}} \tag{5.5}$$

Filter used for the froth dielectric model:

$$\frac{0.0303}{1.000 - 1.2385q^{-1} + 0.6227q^{-2} - 0.3539q^{-3}} \tag{5.6}$$

Filter used for the middling density model:

$$\frac{0.0430}{1.000 - 0.9874q^{-1} + 0.3093q^{-2} - 0.2790q^{-3}} \tag{5.7}$$

One of the advantages mentioned earlier of the PLS technique is it's ability to reduce a problem with a high dimensionality to a lower dimension one. In order to decide how many latent variables to retain for the final model development, the PLS analysis on the filtered data was carried and all eleven latent variables were included. Tables 5-12 to 5-15 give the PLS result for all latent variables for the four models and Figure 5-12 is a plot of the corresponding regression weights.

Latent Variable	Input Block		Output Block	
Number	Individual	Total	Individual	Total
1	46.6814	46.6814	57.6465	57.6465
2	20.5001	67.1815	13.2469	70.8934
3	15.9398	83.1212	2.5052	73.3986
4	3.5579	86.6791	3.9758	77.3743
5	1.7728	88.4519	2.1282	79.5025
6	1.8894	90.3413	0.5981	80.1006
7	2.7966	92.9379	0.2588	80.3594
8	5.3574	98.2952	0.0503	80.4097
9	1.2656	99.5608	0.0546	80.4643
10	0.4350	99.9958	0.1169	80.5812
11	0.0042	100.0000	1.7889	82.3701

Table 5-12: Percent variance captured by the PLS model for froth density when all eleven latent variables are retained (MISO approach).

Latent Variable Number	Input Block		Output Block	
	Individual	Total	Individual	Total
1	50.5741	50.5741	50.9952	50.9952
2	11.6047	62.1787	9.8471	60.8422
3	18.1945	80.3732	5.5241	66.3664
4	7.3385	87.7117	5.7200	72.0863
5	8.7525	96.4643	0.8895	72.9758
6	1.9024	98.3667	0.7964	73.7722
7	0.9674	99.3341	0.4142	74.1863
8	0.5129	99.8469	0.1101	74.2964
9	0.1093	99.9562	0.4353	74.7317
10	0.0239	99.9801	3.8940	78.6258
11	0.0199	100.0000	2.8715	81.4973

Table 5-13: Percent variance captured by the PLS model for froth flow rate when all eleven latent variables are retained (MISO approach).

Latent Variable Number	Input Block		Output Block	
	Individual	Total	Individual	Total
1	46.4623	46.4623	62.2866	62.2866
2	25.7303	72.1926	11.8799	74.1665
3	12.4790	84.6716	4.0678	78.2343
4	5.9173	90.5889	3.6055	81.8398
5	4.9004	95.4893	2.2371	84.0769
6	1.4925	96.9818	2.0266	86.1035
7	1.3741	98.3559	0.1322	86.2357
8	0.9804	99.3363	0.1054	86.3410
9	0.6031	99.9394	0.0577	86.3987
10	0.0581	99.9976	0.0026	86.4013
11	0.0024	100.0000	0.0239	86.4251

Table 5-14: Percent variance captured by the PLS model for froth dielectric when all eleven latent variables are retained (MISO approach).

Latent Variable Number	Input Block		Output Block	
	Individual	Total	Individual	Total
1	49.6594	49.6594	65.4732	65.4732
2	13.1792	62.8386	12.5529	78.0260
3	14.8571	77.6957	3.5650	81.5910
4	10.9067	88.6024	3.2012	84.7922
5	6.2256	94.8280	0.3222	85.1144
6	2.3957	97.2237	0.2910	85.4054
7	1.2802	98.5039	0.2790	85.6845
8	0.6558	99.1597	0.1648	85.8493
9	0.7395	99.8992	0.0109	85.8602
10	0.0768	99.9760	0.2256	86.0858
11	0.0240	100.0000	0.4484	86.5342

Table 5-15: Percent variance captured by the PLS model for middling density when all eleven latent variables are retained (MISO approach).

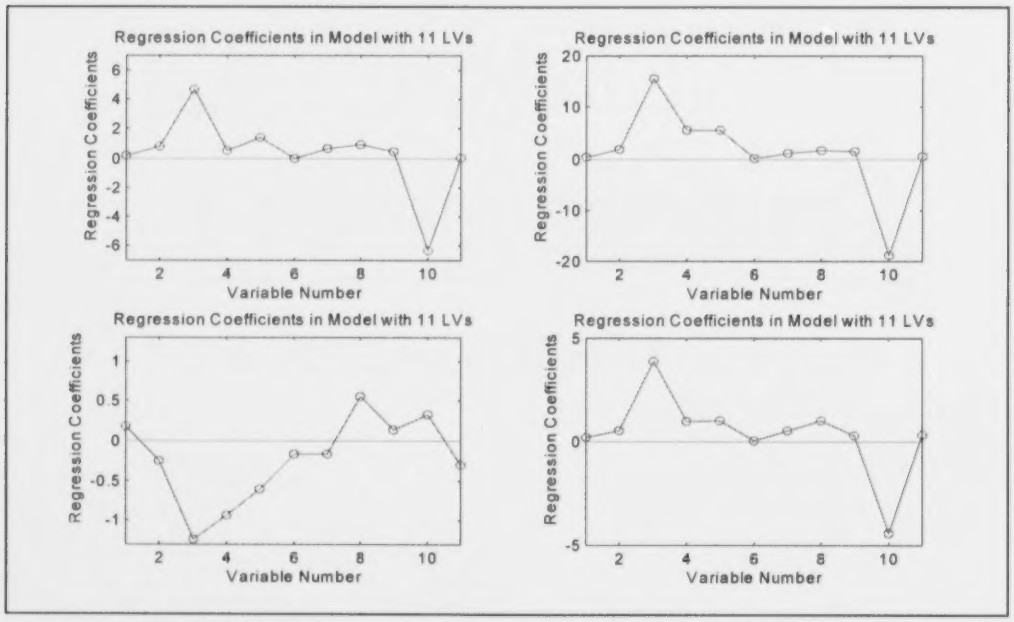

Figure 5-12: Plot of regression coefficients (MISO approach) for eleven latent variable model for froth density (top, left), froth flow rate (top, right), froth dielectric (bottom, left) and middling density (bottom, right).

Examination of the above data indicates, that many of the latent variables do not significantly contribute to explaining the variance observed in the outputs. Therefore, for the four models developed here, between four and six latent variables will be used, depending on which output is modeled. Using more latent variables will result in overfitting the problem and modeling of noise which will lead to reduced overall predictive capabilities for the models. Tables 5-16 to 5-19 summarize the PLS results using only the minimum number of latent variables and Figure 5-13 is a plot of the regression weights for these models. To arrive at the best possible parameter estimates for the models, cross validation was used during development.[1] The data was split into a validation set of 250 data points and the rest of the data was used for identification. The model was rebuilt and tested 10 times with randomly selected validation sets. The final regression vectors are presented in Table 5-20.

Latent Variable	Input Block		Output Block	
Number	Individual	Total	Individual	Total
1	46.6814	46.6814	57.6465	57.6465
2	20.5001	67.1815	13.2469	70.8934
3	15.9398	83.1212	2.5052	73.3986
4	3.5579	86.6791	3.9758	77.3743
5	1.7728	88.4519	2.1282	79.5025

Table 5-16: Percent variance captured by the PLS model for froth density when five latent variables are retained (MISO approach).

Latent Variable	Input Block		Output Block	
Number	Individual	Total	Individual	Total
1	50.5741	50.5741	50.9952	50.9952
2	11.6047	62.1787	9.8471	60.8422
3	18.1945	80.3732	5.5241	66.3664
4	7.3385	87.7117	5.7200	72.0863

Table 5-17: Percent variance captured by the PLS model for froth flow rate when four latent vectors are retained (MISO approach).

Latent Variable	Input Block		Output Block	
Number	Individual	Total	Individual	Total
1	46.4623	46.4623	62.2866	62.2866
2	25.7303	72.1923	11.8799	74.1665
3	12.4790	84.6716	4.0678	78.2343
4	5.9173	90.5889	3.6055	81.8398
5	4.9004	95.4893	2.2371	84.0769
6	1.4925	96.9818	2.0266	86.1035

Table 5-18: Percent variance captured by the PLS model for froth dielectric when six latent variables are retained (MISO approach).

Latent Variable	Input Block		Output Block	
Number	Individual	Total	Individual	Total
1	49.6594	49.6594	65.4732	65.4732
2	13.1792	62.8386	12.5529	78.0260
3	14.8571	77.6957	3.5650	81.5910
4	10.9067	88.6024	3.2012	84.7922

Table 5-19: Percent variance captured by the PLS model for middling density when four latent variables are retained (MISO approach).

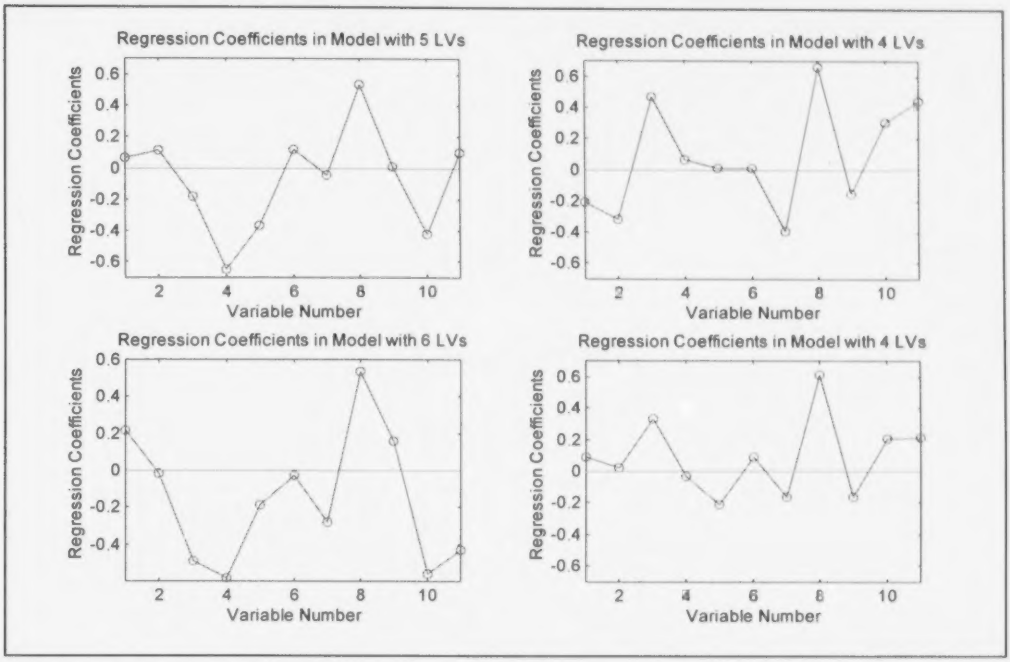
Regression Coefficients in Model with 5 LVs
Regression Coefficients in Model with 4 LVs
Regression Coefficients in Model with 6 LVs
Regression Coefficients in Model with 4 LVs
Regression Coefficients
Variable Number

Figure 5-13: Plot of regression coefficients (MISO approach) for reduced number of latent variable model for froth density (top, left), froth flow rate (top, right), froth dielectric (bottom, left) and middling density (bottom, right).

Variable Number	Measurement	Froth Density	Froth Flow	Froth Quality	Middling Density
1	Oil sand gamma count	0.0632	-0.2107	0.2168	0.0907
2	Oil sand bitumen content	0.1138	-0.3169	-0.0168	0.0200
3	PSV feed density	-0.1799	0.4704	-0.4916	0.3352
4	Tumbler water flow rate	-0.6447	0.0676	-0.5801	-0.0339
5	Flood water flow rate	-0.3650	0.0121	-0.1871	-0.2111
6	Caustic flow rate	0.1197	0.0084	-0.0287	0.0887
7	Tumbler steam flow rate	-0.0430	-0.3940	-0.2777	-0.1615
8	Reject rate	0.5365	0.6577	0.5381	0.6101
9	Oil sand water content	0.0122	-0.1571	0.1588	-0.1601
10	Oil sand feed rate	-0.4177	0.3043	-0.5571	0.2013
11	PSV interface level	0.1003	0.4413	-0.4339	0.2103

Table 5-20: Normalized regression coefficients for the PLS models (MISO approach).

The relative magnitude of the regression coefficients indicate how influential a particular input parameter is on the output of interest. The operational implications are that the operator now has a tool which assists him/her in deciding which inputs to change to achieve a given operating goal. For example, if the froth density must be adjusted, based on the results presented in Table 5-20, the most effective way would be to manipulate the tumbler water flow rate. Additionally, the sign of the regression coefficient indicates if a particular input is directly or inversely related to a given output. However, it is also clear that this is not the only way to affect the froth density. Since the results identify several parameters which could be used to influence a given output, the operator can achieve the same goal even though one of the variables might be constrained for some reason. Continuing with the froth density example, if the tumbler water can not be changed due to hot water limitations, adjusting the oil sand feed rate is almost as effective.

5.5.2. Single PLS model (MIMO approach)

Again both the input and output data was filtered prior to model identification. However, since the filters identified previously were tailored to each output rather than to the system as a whole, there is no basis for selecting any one of them in this case. The filter chosen for this investigation was a simple first order filter with a time constant similar to the one for the slowest filter for the MISO approach. The filter used for the MIMO model is given by:

$$\frac{0.02}{1.00 - 0.98q^{-1}} \tag{5.8}$$

In order to decide how many latent variables to retain for the final model development, the PLS analysis on the filtered data was carried and all eleven latent

Latent Variable Number	Input Block		Output Block	
	Individual	Total	Individual	Total
1	49.9986	49.9986	52.7813	52.7813
2	23.8583	73.8569	16.8392	69.6224
3	13.0577	86.9146	4.7760	74.3984
4	5.5743	92.4889	4.5292	78.9276
5	4.0435	96.5324	2.8874	81.8150
6	1.7311	98.2635	2.0347	83.8497
7	0.9992	33.2627	1.3479	85.1976
8	0.5649	99.8277	0.3351	85.5327
9	0.1313	99.9589	0.2082	85.7409
10	0.0246	99.9835	1.0163	86.7573
11	0.0165	100.0000	1.3707	88.1280

Table 5-21: Percent variance captured by the PLS model when all eleven latent variables are retained (MIMO approach).

variables were included. Table 5-21 gives the PLS result for all latent variables for the model and Figure 5-14 is a plot of the corresponding regression weights.

As expected, many of the latent variables did not significantly contribute to explaining the variance observed in the outputs. Six latent variables, explaining about 84% of the total output variance, will be used for the final MIMO model to prevent modeling of noise and overfitting of the data set. Table 5-22 summarizes the PLS results using only the six most significant latent variables. Figure 5-15 is a plot of the regression weights obtained for this model. The procedure employed for this case was identical to the one used for the MISO approach. The data was split into a 250 point validation set with the rest being used for model identification during cross validation. The model was rebuilt and tested 10 times with randomly selected validation sets. The final regression vectors are presented in Table 5-23.

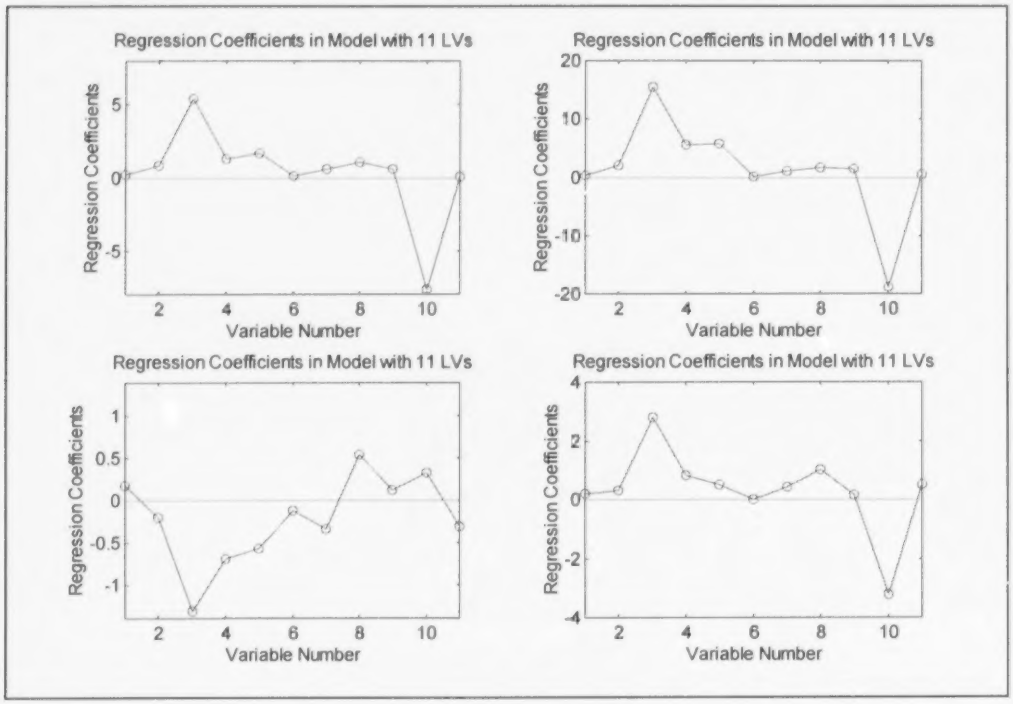

Figure 5-14: Plot of regression coefficients (MIMO approach) for eleven latent variable model for froth density (top, left), froth flow rate (top, right), froth dielectric (bottom, left) and middling density (bottom, right).

Latent Variable Number	Input Block		Output Block	
	Individual	Total	Individual	Total
1	49.9986	49.9986	52.7831	52.7831
2	23.8583	73.8569	16.8392	69.6224
3	13.0577	86.9146	4.7760	74.3984
4	5.5743	92.4889	4.5292	78.9276
5	4.0435	96.5324	2.8874	81.8150
6	1.7311	98.2635	2.0347	83.8497

Table 5-22: Percent variance captured by the PLS model when six latent variables are retained (MIMO approach).

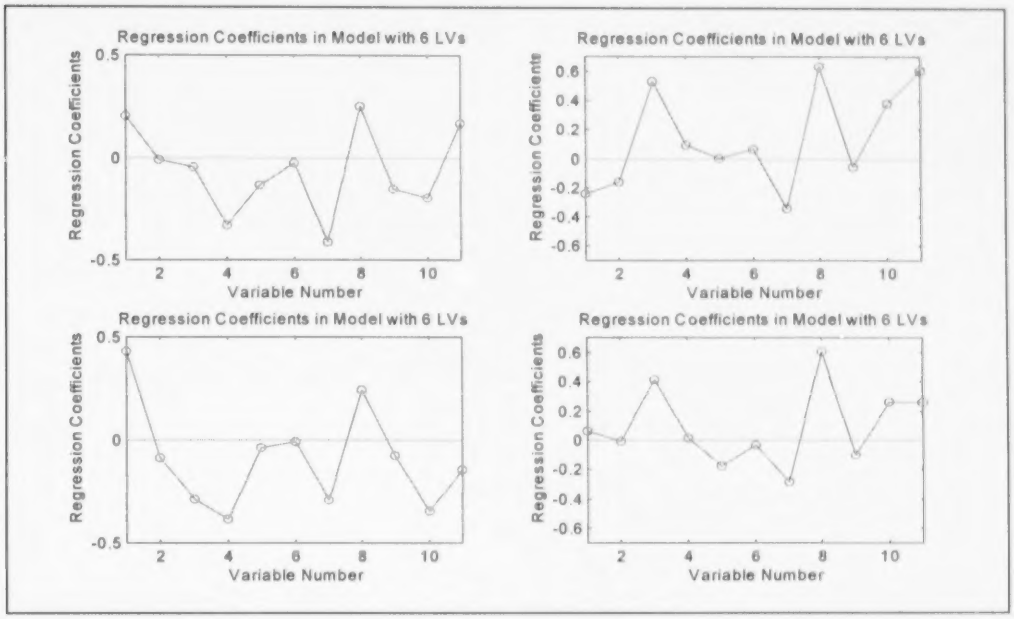

Figure 5-15: Plot of regression coefficients (MIMO approach) for six latent variable model for froth density (top, left), froth flow rate (top, right), froth dielectric (bottom, left) and middling density (bottom, right).

Variable Number	Measurement	Froth Density	Froth Flow	Froth Quality	Middling Density
1	Oil sand gamma count	0.2032	-0.2444	0.4318	0.0598
2	Oil sand bitumen content	-0.0141	-0.1630	-0.0862	-0.0082
3	PSV feed density	-0.0465	0.5270	-0.2851	0.4150
4	Tumbler water flow rate	-0.3270	0.0892	-0.3844	0.0127
5	Flood water flow rate	-0.1364	0.0013	-0.0399	-0.1760
6	Caustic flow rate	-0.0278	0.0619	-0.0115	-0.0360
7	Tumbler steam flow rate	-0.4100	-0.3426	-0.2913	-0.2876
8	Reject rate	0.2503	0.6340	0.2408	0.6057
9	Oil sand water content	-0.1559	-0.0605	-0.0764	-0.0966
10	Oil sand feed rate	-0.1969	0.3779	-0.3428	0.2590
11	PSV interface level	0.1676	0.6037	-0.1459	0.2597

Table 5-23: Normalized regression coefficients for the PLS models (MIMO approach).

5.6. PLS model validation

Proper model validation requires the use of a different data set for testing than was used for development. Since the data set for September 5th was used to develop the model, it was verified by comparing the model predictions for the other four days of data to the actual outputs. Prior to developing the PLS model, the identification data set was mean centered and scaled to unit variance. Therefore, the data used for the validation must also be normalized by subtracting the same mean and dividing by the same variance that were used in the normalization of the model identification data set. Also, the data filters developed for the identification data set were applied to the validation data sets. The regression coefficients calculated during model development were then used to determine the model outputs:

$$\hat{y} = \beta_1 u_1 + \beta_2 u_2 + \beta_3 u_3 + \ldots + \beta_{11} u_{11} \tag{5.9}$$

where: $\hat{y}$ = predicted output

β_n = nth regression coefficient

u_n = nth input

In order to check that the models developed are at least self consistent, the predicted outputs for the September 5th data set (used for model identification) were also included in the validation.

5.6.1. Validation for the MISO approach

Figures 5-16 to 5-20 are plots of predicted versus actual outputs using the regression weights obtained by the MISO approach for the four outputs. Table 5-24 lists the mean value for the residuals for the model predictions. These indicate the average amount of deviation between the actual outputs and the model predictions. The expected value for them is zero.

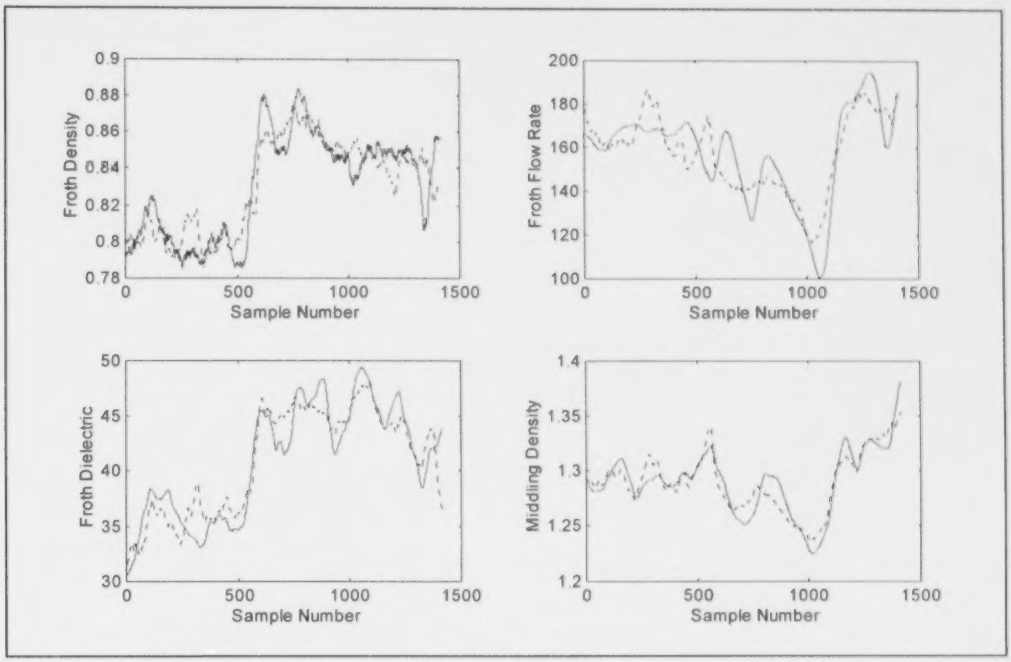

Figure 5-16: Comparison of actual measurements (—) to model predictions (---) for September 5th data set (MISO approach).

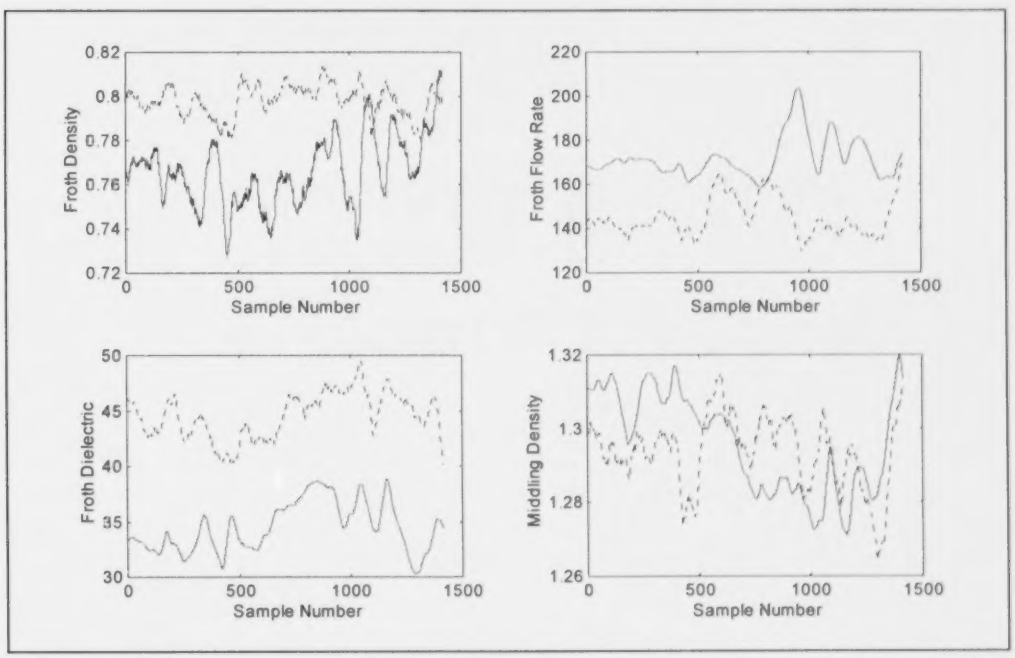

Figure 5-17: Comparison of actual measurements (—) to model predictions (---) for September 4th data set (MISO approach).

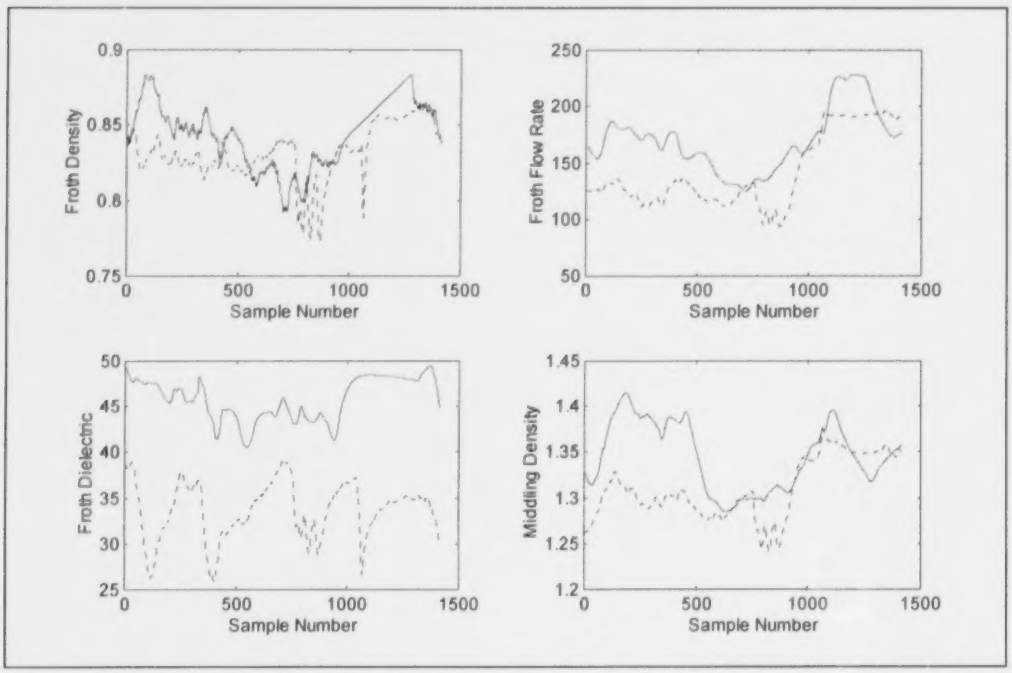

Figure 5-18: Comparison of actual measurements (—) to model predictions (---) for September 6th data set (MISO approach).

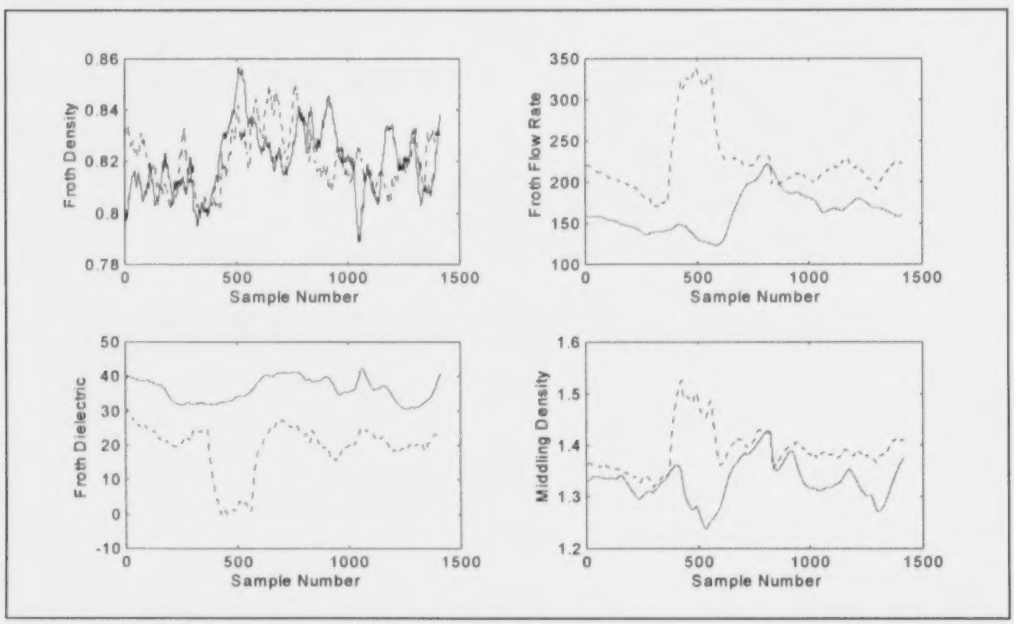

Figure 5-19: Comparison of actual measurements (—) to model predictions (---) for September 7th data set (MISO approach).

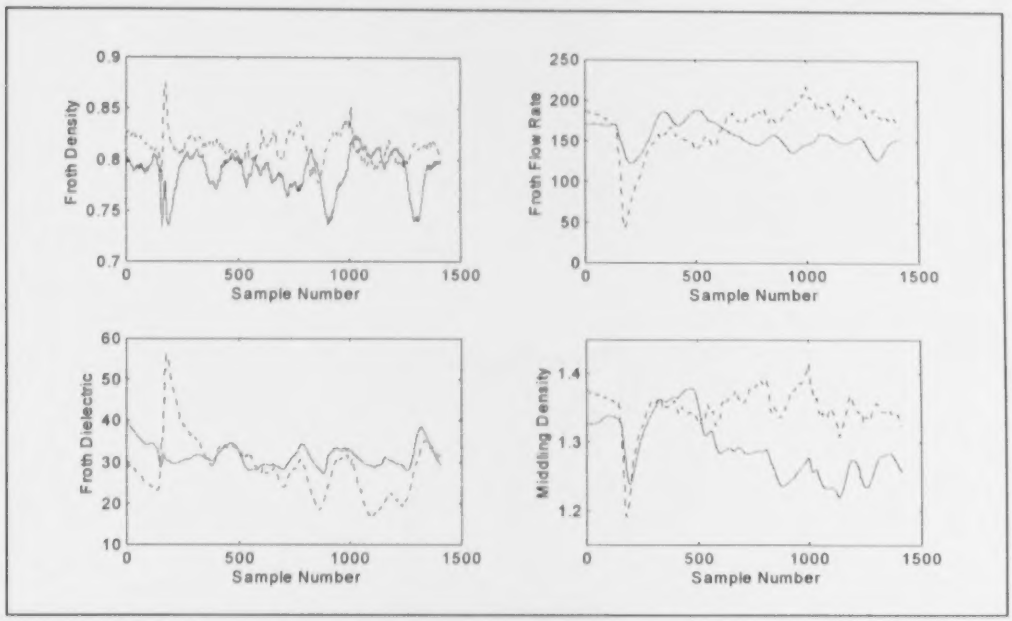

Figure 5-20: Comparison of actual measurements (—) to model predictions (---) for September 9th data set (MISO approach).

As seen in Figure 5-16, the predictions of the model for the day it has been developed are, for the most part, quite reasonable. The model captures the changes in the magnitude of the measurements but does not predict the dynamics as well. However, with a few exceptions, the predictive capability of the model for the other days that it was tested on, is poor. This finding is consistent with the observations Section 5.3 that showed large differences in the total percent variance captured by the model from one day to the next (Tables 5-6 to 5-9).

Measurement	Sept. 4th	Sept. 5th	Sept. 6th	Sept. 7th	Sept. 9th
Froth density	-4.21	8×10^{-14}	1.47	-0.11	-3.10
Froth flow rate	15.69	-2×10^{-13}	16.63	-37.27	-7.74
Froth dielectric	-29.14	-7×10^{-13}	26.54	46.10	8.14
Middling density	0.18	-1×10^{-13}	2.59	-4.60	-4.39

Table 5-24: Relative mean of residuals (%) when comparing the model output to actual measurements (MISO approach).

5.6.2. Validation for the MIMO approach

Similar to the validation performed for the MISO approach, Figures 5-21 to 5-25 are plots of predicted versus actual outputs for the MIMO approach.. Table 5-25 lists the mean value for the residuals for the model predictions.

Again, as for the MISO approach, the model gives good predictions for the data that was used during the development, but for the most part, other data sets show large discrepancies between the predicted and measured values. Also a visual comparison of the predictions obtained by the two methods show significant differences. From the data sets examined here, it is not possible to draw any definite conclusions as to which of the two approaches is better suited.

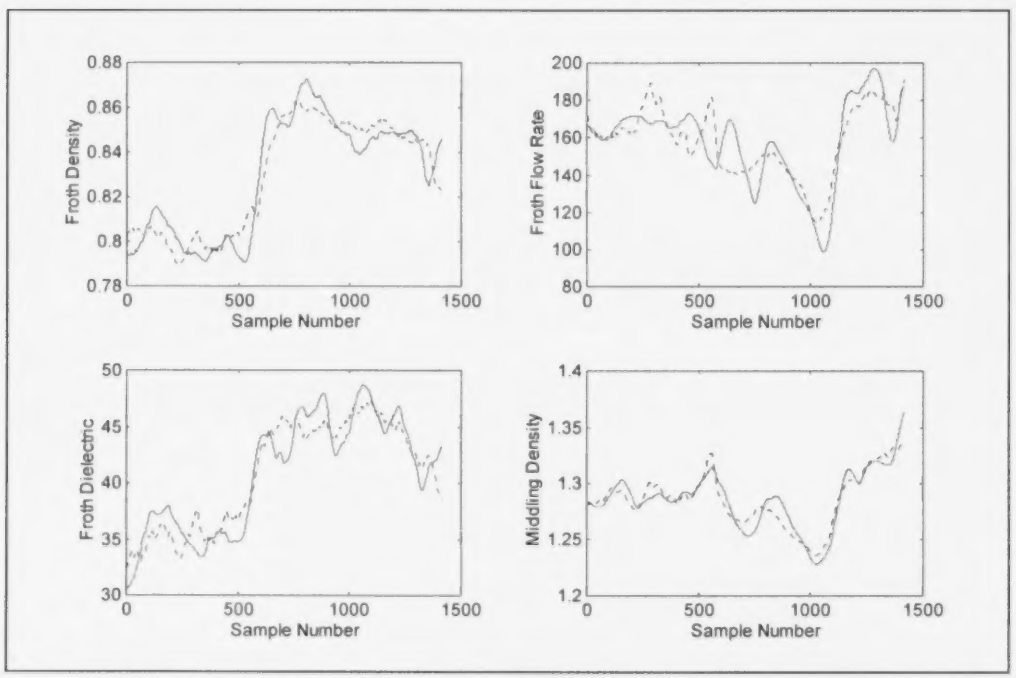

Figure 5-21: Comparison of actual measurements (—) to model predictions (---) for September 5th data set (MIMO approach).

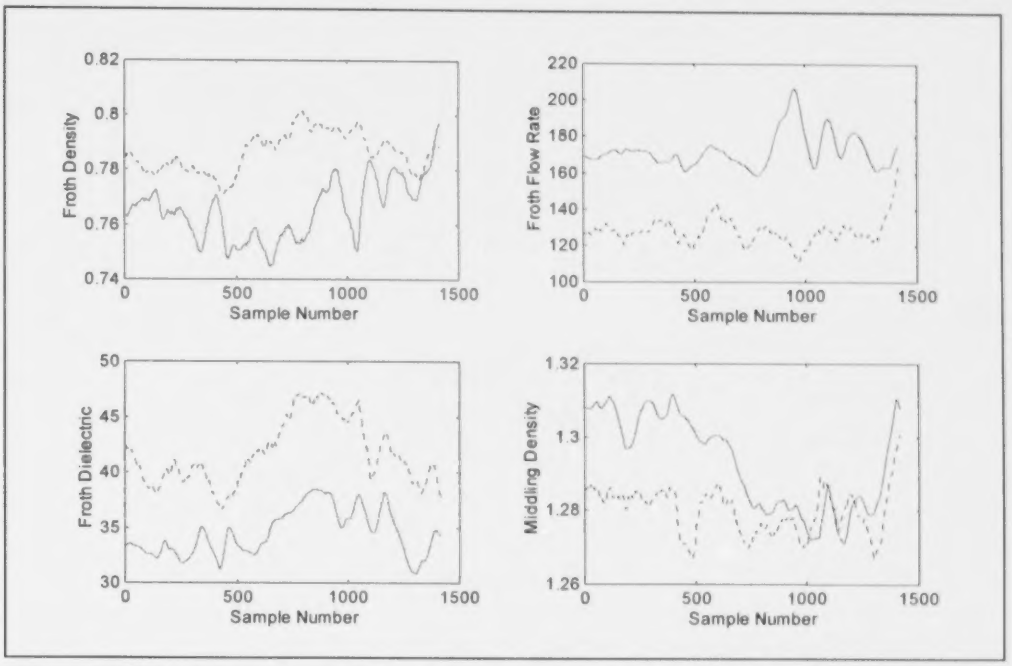

Figure 5-22: Comparison of actual measurements (—) to model predictions (---) for September 4th data set (MIMO approach).

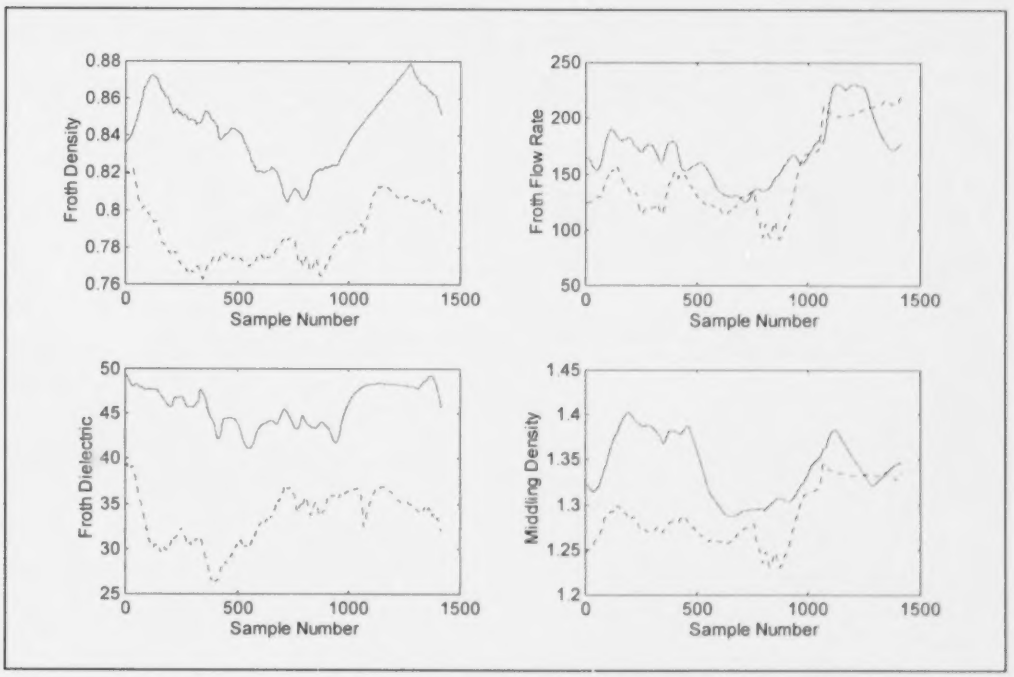

Figure 5-23: Comparison of actual measurements (—) to model predictions (---) for September 6th data set (MIMO approach).

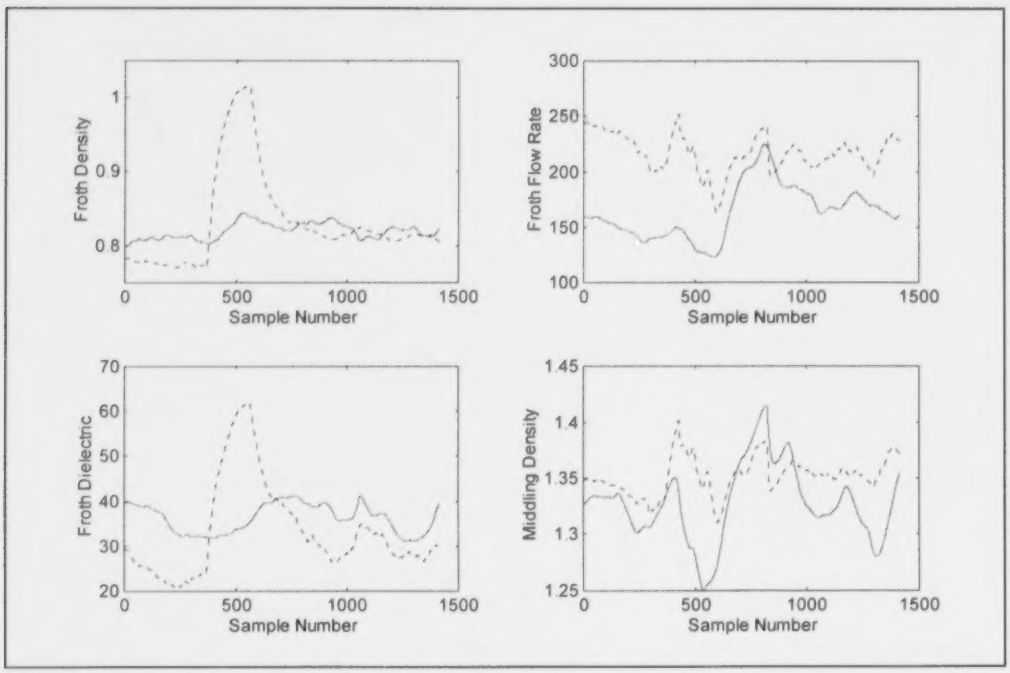

Figure 5-24: Comparison of actual measurements (—) to model predictions (---) for September 7th data set (MIMO approach).

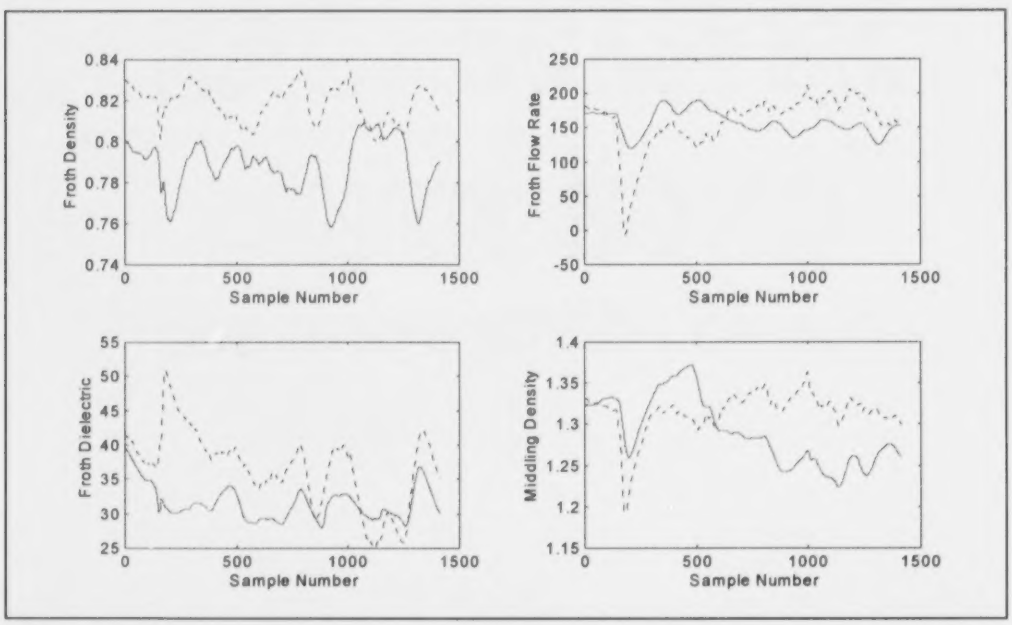

Figure 5-25: Comparison of actual measurements (—) to model predictions (---) for September 9th data set (MIMO approach).

Measurement	Sept. 4th	Sept. 5th	Sept. 6th	Sept. 7th	Sept. 9th
Froth density	-2.80	5×10^{-8}	6.76	-1.65	-3.90
Froth flow rate	25.81	1×10^{-10}	11.96	-32.19	-0.74
Froth dielectric	-20.90	-6×10^{-9}	27.13	8.00	-16.25
Middling density	0.99	2×10^{-9}	3.96	-1.84	-1.91

Table 5-25: Relative mean of residuals (%) when comparing the model output to actual measurements (MIMO approach).

5.7. Adaptive model

The most obvious shortcoming of the PLS model described above is that, in most cases, it exhibits a large bias for all data sets except the one that was used during model development. The most probable explanations for this are changes in operating modes (periodic operation of recycle loops, for example) in addition to other disturbances. Since neither of these are recorded in the data base, it is impossible to incorporate them into the model. One way to improve the model performance under such conditions is to make it adaptive. Presented in this section is an adaptive model for the PSV.

The fixed model used for the simulations for the MISO case (Section 5.6.1) is given by:

$$\hat{y} = \beta_1 u_1 + \beta_2 u_2 + \ldots + \beta_{11} u_{11} \tag{5.10}$$

where: $\hat{y}$ = predicted output
β_n = regression coefficient for n^{th} input
u_n = n^{th} input

The proposed adaptive model is based on the above equation, but contains an additional factor that can be dynamically adjusted to compensate for any unrecorded changes in the process. The adaptive model equation is:

$$\hat{y} = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{11} x_{11} + \lambda \qquad (5.11)$$

where: $\hat{y}$ = predicted output
β_n = regression coefficient for n^{th} input
x_n = n^{th} input
λ = model update parameter

The model update parameter in Equation 5.11 can be viewed as a model bias correction term that is modified periodically based on past model - process mismatch. This update method relies on the fact that many of the disturbances the model is adapting to are slow compared to the length of the parameter update window. In the case of changes to the plant operating modes, this is definitely the case. The status of recycle streams, for example, are changed in the order of once or twice per day at the most. However, for any disturbances that affect the output within a time frame close to, or faster than the length of the update window the update mechanism proposed here will not be adequate.

5.7.1. Simulation results using the adaptive model

The simulation carried out to test the performance of the adaptive model were for the MISO case using the September 5^{th} , 6^{th} and 7^{th} data sets. The model update parameter was adjusted once every 60 samples (i.e. once per hour of process time), resulting in a prediction horizon for the model of 1 hour. The results of simulations for the adaptive model are given in Figures 5-26 to 5-28.

Since the fixed model for the MIMO approach is almost identical to the one for the MISO approach, the adaptive model was not tested for the MIMO case, even though it could quite easily be extended for it. It is expected that the results would be almost identical to the ones presented here.

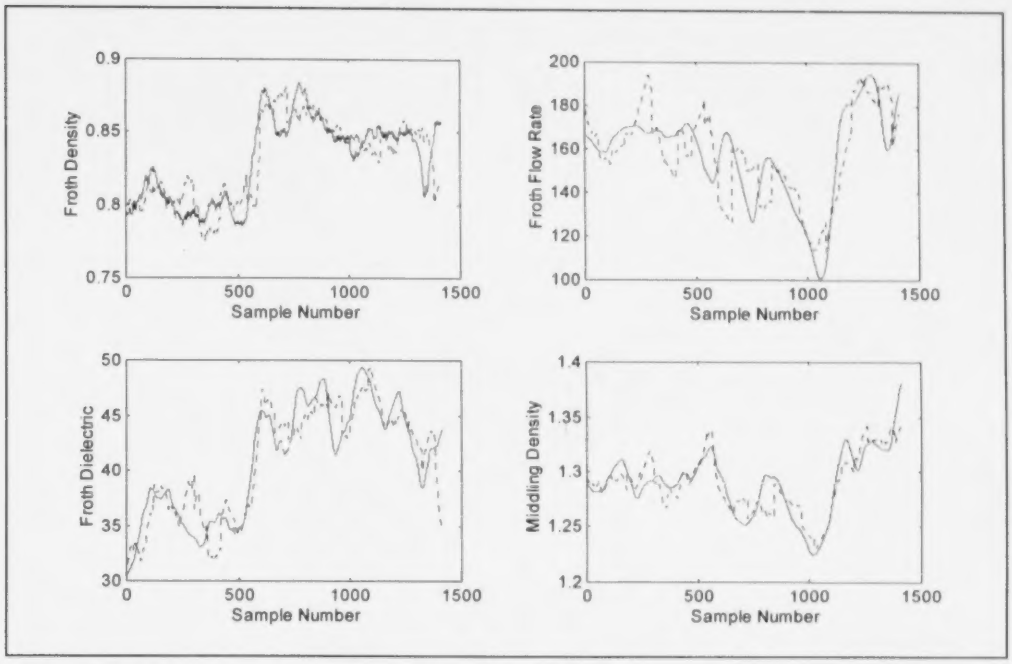

Figure 5-26: Comparison of actual measurements (—) to model predictions (---) for September 5th data set using the adaptive model.

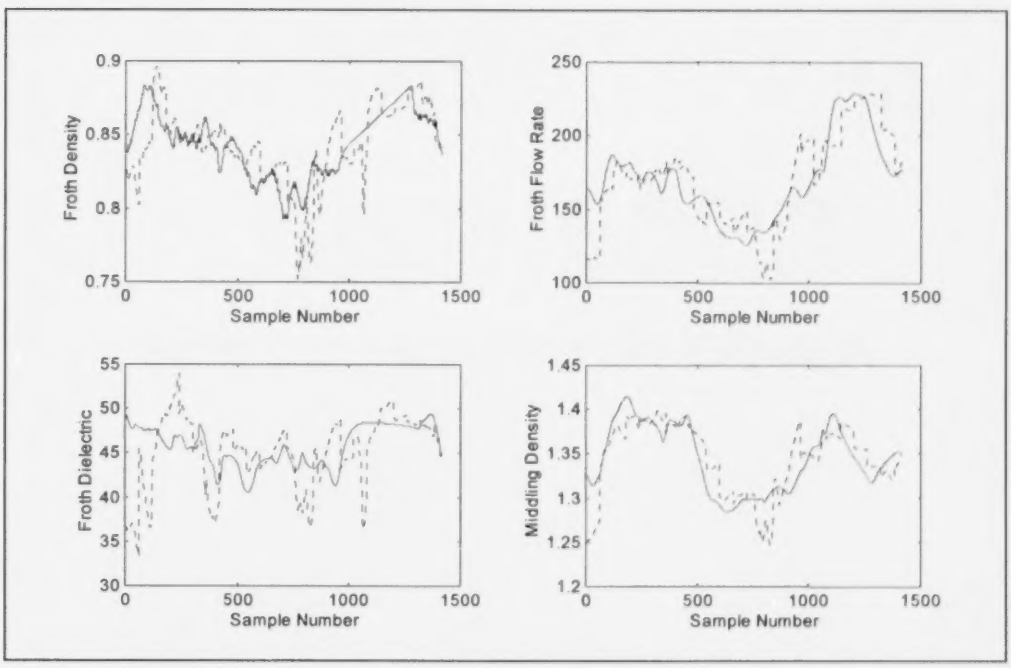

Figure 5-27: Comparison of actual measurements (—) to model predictions (---) for September 6th data set using the adaptive model.

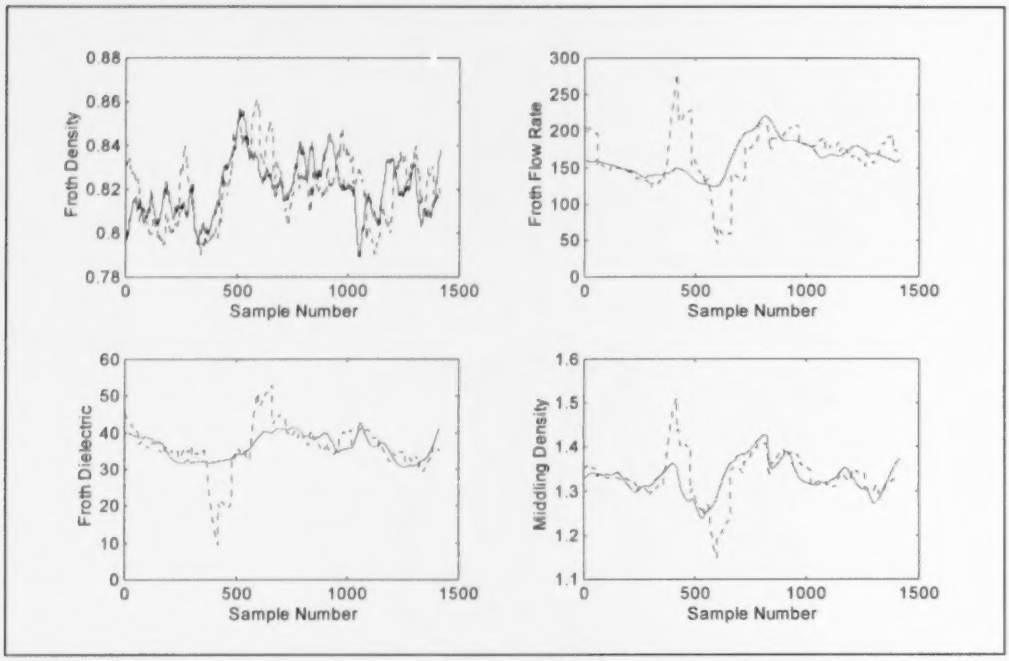

Figure 5-28: Comparison of actual measurements (—) to model predictions (---) for September 7th data set using the adaptive model.

One important point to note is that in the case were the fixed model gives good predictions, the adaptive model does not have a negative impact on the results (Figures 5-26 and 5-16). For the data sets where the fixed model exhibited large offsets, the adaptive model improved the agreement with the plant output significantly (Figures 5-27 and 5-18 for the September 6th data set and Figures 5-28 and 5-19 for September 7th). While this new model was quite effective in removing the large observed biases, the model has difficulty at times matching the dynamics of the process. There are several reasons for this behavior:

- the model is a linear approximation of a non-linear process
- significant, unmeasured process disturbances occur too fast for the adaptive algorithm to compensate for

- unreconciled input and output data

Overall, while the adaptive model still falls short of giving reliable enough results for on-line implementation, it is a significant improvement over the fixed model. One possible way to further improve the adaptive model is by fine tuning λ between the 60 minute update intervals based on some knowledge of the expected trajectory of the parameter. Figures 5-29 to 5-31 show the values for λ for the data presented here. Unfortunately, the values do not exhibit any discernible pattern.

5.8. Conclusion

The investigations carried out here showed that the PLS method is quite sensitive to errors in the process dead time as well as to the lack of steady-state conditions. This poses a significant problem for the development of a PLS model for

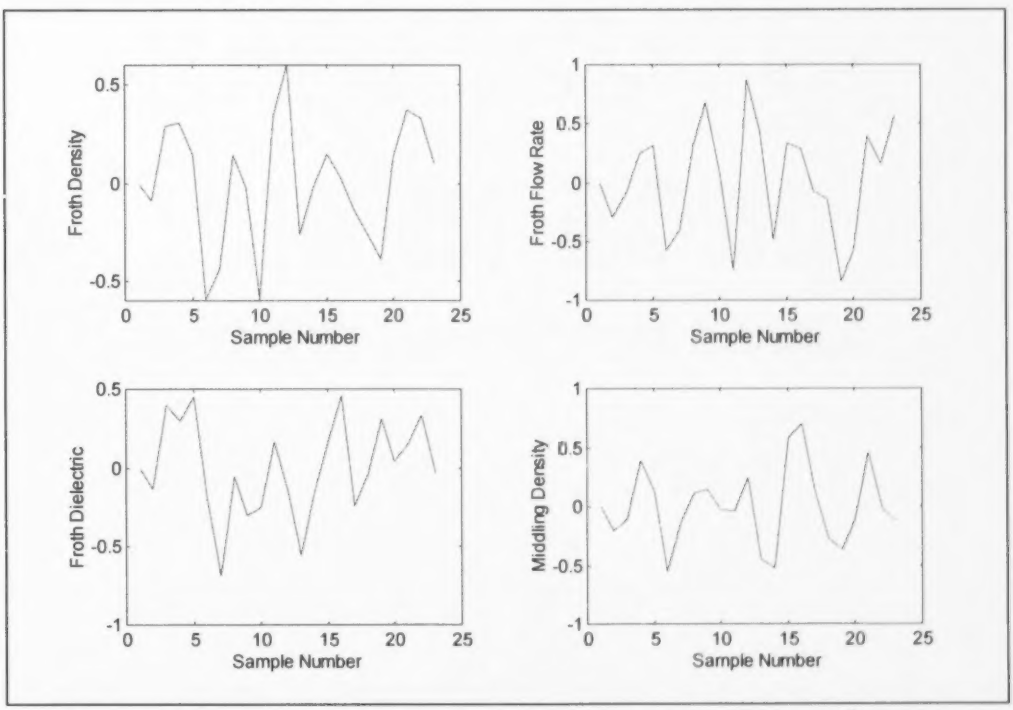

Figure 5-29: Trajectory of model update parameter for September 5th data set.

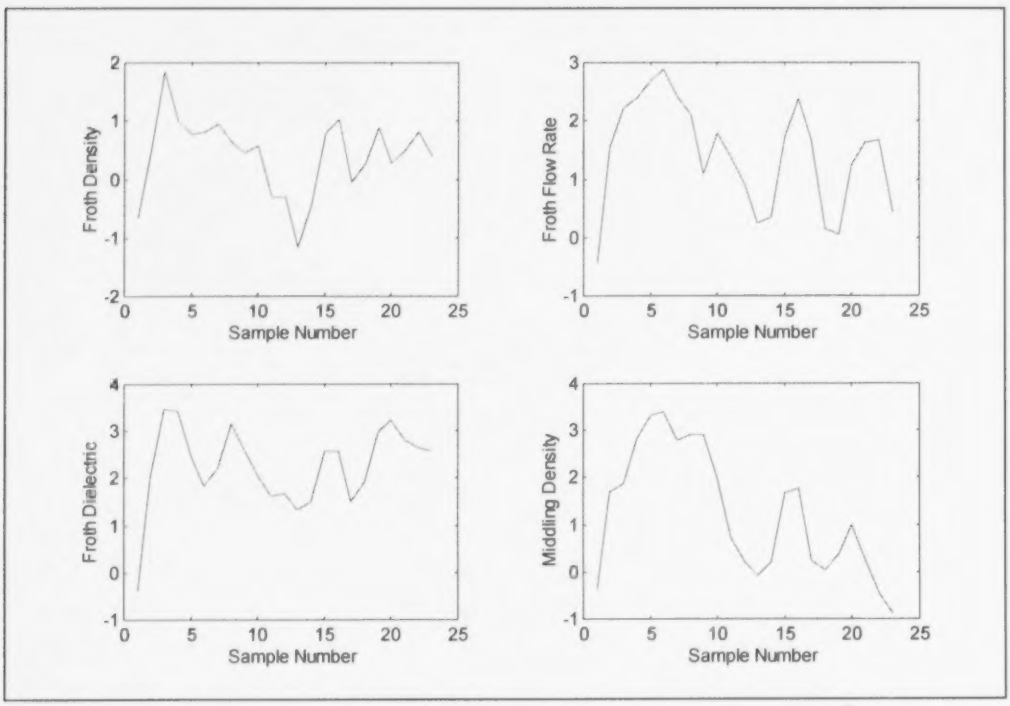

Figure 5-30: Trajectory of model update parameter for September 6th data set.

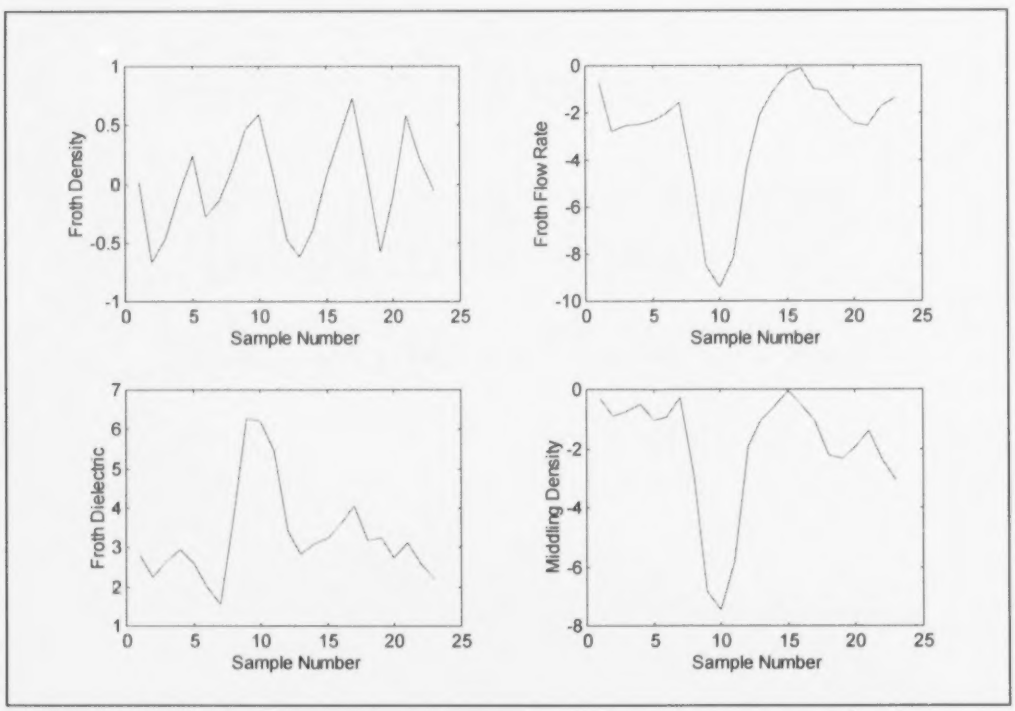

Figure 5-31: Trajectory of model update parameter for September 7th data set.

the extraction plant. Many of the time delays in the system are a function of the amount of oil sand processed, the volume of water added and the different operating conditions chosen by the operators. As was found in Chapter 4, some of the delay times can vary by up to ±6 minutes. This amount of error in the delay time will seriously affect the model's performance.

It was found, that the model developed using the September 5th data set did capture a significant amount of the output variation for this data set. This is illustrated by the relatively good agreement between the actual and the predicted outputs as well as by the low bias in the residuals for this day. It is equally clear, however, that this model does not predict the system behavior for the other data sets on which it was tested. It is encouraging however, that in a few cases, similarities in overall trends can be seen, even though large offsets are present. One such example is the froth dielectric predicted by the MIMO model for September 4th. Another one is the froth density predictions (MISO model) for the September 7th data set. Here, the model does not predict the short term variations very well, but the level predictions are relatively close. The differences observed appear to be too large though to be solely attributable to changes in the delay times. Unmeasured disturbances and unrecorded changes in operating conditions contribute as well to a large extend.

The adaptive model proposed in Section 5.7 addresses some of these concerns. It was quite effective in removing biases caused by infrequent changes in operating conditions. The simulations for the adaptive model however, show quite clearly that a linear approximation of the process does not yield satisfactory results. A non-linear, adaptive model that includes some form of real time data validation and reconciliation is needed.

Differences between the multiple MISO and single MIMO approaches were

also observed. The four separate MISO models gave generally better predictions (based on prediction bias) than the single MIMO approach. This is somewhat surprising since comparing the total output variance explained by the two approaches are quite similar: 80%, 72%, 86% and 85% for the four MISO models and 84% for the MIMO model. Comparing the four regression vectors for the two approaches shows significant differences between their magnitudes and in some cases, even the sign is different. On the other hand, a comparison between Figures 5-13 and 5-15 shows that the distribution pattern for the regression weights for the two approaches are similar and that for both cases, the same inputs are identified as significant.

The heavy filtering that was necessary for these data sets makes it impossible for this approach to meet the original objectives, even if the models were capable of perfectly predicting the outputs. Taking the filter time constants into account, the time resolution of the models is in the order of one hour or more and not the two to five minutes required for an operating tool.

The findings here do not rule out the usefulness of PLS for building a model for the extraction process, but rather indicate that a significant amount of work is required in improving the current process measurements, adding important process parameters that are not in the present data base, optimizing the existing regulatory control loops and developing good on-line, real time data validation techniques.

5.9. References

[1] Wold, S., "Cross-Validatory Estimation of the Number of Components in Factor and Principal Component Models", Technometrics 20, 397-405 (1978).

[2] Wold, H., "Soft Modeling. The Basis Design and Some Extensions". Systems Under Indirect Observation, K. Joreskog and H. Wold, Ed., North Holland

Amsterdam (1982).

[3] Wold, S., A. Ruhe, H. Wold and W. Dunn III, "The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses", SIAM J. Sci. Stat. Comput. 5, 735-743 (1984).

[4] Kresta, J.V., "The Application of Partial Least Squares to Problems in Chemical Engineering", Ph.D. Thesis, McMaster University (1992).

6. Conclusions and Future Work

A truly overwhelming amount of operating data from Syncrude's extraction plant is being collected and archived on a daily basis. Unfortunately however, this enormous source of information remains underutilized to a great extend. The common theme that evolved from the investigations presented here is that some effort needs to be made to improve the quality of the data this database as well as the basic regulatory control of the process.

The ultimate objective of developing a real time model that will assist the operator in maximizing the product quality and quantity from this process and minimizing the cost of operation can only be realized if reliable, high quality data is available for model development, all basic control loops are functioning optimally and all relevant measurements are available and as accurate as possible.

Many of the regulatory control loops have a very fast response time compared to the time constant of the process they are designed to control. This results in the controller responding to random measurement noise in addition to true process variations. The effect of this is twofold: excessive wear on the control element and increased downstream variability. A good example of this is the tumbler water control loop described in Chapter 3 where the application of a simple filter was able to reduce the control effort to one third.

An important aspect to improving the quality of the data collected from the process is real time data validation and reconciliation. Some sensors currently operating in the plant show excessive bias and uncertainty (the oil sand feed rate measurement comes to mind). On-line verification of measurements and model based on-line calibration can significantly improve the quality of the measurements, which

in turn will lead to better control of the process and enable the identification of models with better predictive capabilities.

The large variations in delay time are inherent in the process and this situation can not be improved upon. The implications of this can be divided into two parts: output prediction and model identification. The effect on model prediction is the easier of the two to deal with. If the actual current delay times for the various parts of the process are known, then the appropriate past input readings can be chosen at every prediction step in order to calculate an output estimate. Therefore, a method must be developed to estimate the process time delays in real time based solely on the process inputs and pass them to a model capable of accepting changing time delays. When building a model on the other hand, the data used must be over a sufficiently long time period to obtain statistical significance. The data set must also include as widely varying operating conditions as possible to ensure that the process dynamics are adequately captured. However, this implies that the dead times within the data set will vary considerably, interfering with the model identification process.

The PLS model identified here for the PSV outputs was capable of capturing a fair amount of the process dynamics and gave reasonable predictions for the data set that was used during the development phase. When this model was applied to other data sets however, large biases and differences in the predicted dynamics were observed. These can only partially be attributed to the difficulties mentioned above. A significant portion of the observed deviations can only be explained through the presence of unmeasured disturbances. Some effort should be made in identifying these and devising a strategy for incorporating them into future models. Also the long time constant filters needed to achieve the predictions presented here makes this model unsuitable for model based control or real time operator assistance.

Overall, the analyses carried out here identified key issues that need attention. On the practical side, real time data verification and reconciliation as well as improvements to the regulatory control loops are of prime importance. On the theoretical side, some method must be found that will result in good predictive models in the presence of significantly varying process dead times.

Appendix

Most of the data processing, simulation and data visualization for this thesis has been carried out in Matlab®.* This appendix is a compilation of all special Matlab® programs which have been used in carrying out the research.

All PLS model development was done using the PLS toolbox version 1.3 developed by Barry M. Wise.[1]

The 'inplot.m' function plots a data series and superimposes a small window containing a subset of the data, giving a more detailed view of a particular region. This program is available, free of charge, from The MathWorks, Inc.'s ftp site. It was used without modification.

```
% function [h]=inplot(xvec,yvec,sran,c1,bds1,c2,bds2,c3,bds3)
%
% This function plots the data (xvec,yvec) and then plots a
% subset in the upper left corner of the screen to give you a
% "picture-in-a-picture".
% Jeff Butera, jvbutera@eos.ncsu.edu, 30 November 95 (with much
% help from Mike Buksas).
%
% Required INPUTS:
%   xvec - vector of x data points
%   yvec - vector of y data points, same length as xvec
%   sran - 2 vector of form [smin,smax] for subset to be plotted.
% Note that (smin,smax) must be a subset of (min(xvec),max(xvec)).
%
% Optional INPUTS:
%   c    - a character, either 'w','p', or 's'
%   bds - 4 vector in form [xmin xmax ymin ymax]
%
%   If c=='p' then bds specifies the axis bounds for the outermost
%     Plot.
%   If c=='s' then bds specifies the axis bounds for the Subplot.
%   If c=='w' then bds specifies the absolute bounds of the Window
%     location.
%
%   The bounds for the window location will be "clipped" so the
%   subplot window does not run off the top or sides of the main
%   plot.
```

* A registered trademark of The MathWorks, Inc.

```
%
% OUTPUTS:
%   h - a 2 vector of handles.  The first entry is the main graph
%       and the second is the subplot graph.
%
% Examples:
%  x=[-500:500]; y=randn(1,1001);
%  inplot(x,y,[-200 -180])
%  inplot(x,y,[200 230],'w',[-400 400 4 10])
%  inplot(x,y,[300 315],'p',[-500 500 -3 8],'w',[-400 400 4 7])
%  inplot(x,y,[180 197],'s',[175 200 -1 1],'w',[-400 400 4 11])

function [h]=inplot(xvec,yvec,sran,c1,bds1,c2,bds2,c3,bds3)
%
ymax=max(yvec);
ymin=min(yvec);
xmax=max(xvec);
xmin=min(xvec);
%
% Figure out if user set window, entire plot or subplot axis, and
% which is which...
%
wbds=[0 0 0 0];
pbds=[0 0 0 0];
sbds=[0 0 0 0];
if (nargin>3),
   if ((c1=='w')|(c1=='W')),
     wbds=bds1;
   elseif ((c1=='p')|(c1=='P')),
     pbds=bds1;
   elseif ((c1=='s')|(c1=='S')),
     sbds=bds1;
   end
end
if (nargin>5),
   if ((c2=='w')|(c2=='W')),
     wbds=bds2;
   elseif ((c2=='p')|(c2=='P')),
     pbds=bds2;
   elseif ((c2=='s')|(c2=='S')),
     sbds=bds2;
   end
end
if (nargin>7),
   if ((c3=='w')|(c3=='W')),
     wbds=bds3;
   elseif ((c3=='p')|(c3=='P')),
     pbds=bds3;
   elseif ((c3=='s')|(c3=='S')),
     sbds=bds3;
   end
end
if (pbds==[0 0 0 0]),
   pbds=[xmin xmax ymin-0.15*(ymax-ymin) ymax+1.25*(ymax-ymin)];
```

```
end
plenx=pbds(2)-pbds(1);
pleny=pbds(4)-pbds(3);
%
% plot data in large window with plot axis already set
%
clf
plot(xvec,yvec)
axis(pbds)
hold on
%
% Check to see if user set window size manually
%
if (wbds==[0 0 0 0]),
   wbds=[0.2 0.6 0.3 0.25];
else
%
% Clip subplot window so it doesn't run off top or sides of figure
%
   tbds=wbds;
   wbds(1)=max(0.775*(tbds(1)-pbds(1))/plenx+0.13,0.2);
   wbds(2)=0.815*(max(tbds(3),pbds(1))-pbds(3))/pleny+0.11;
   wbds(3)=min((0.775*(tbds(2)-pbds(1))/plenx)+0.13-wbds(1),0.85-
wbds(1));
   wbds(4)=min((0.815*(tbds(4)-pbds(3))/pleny)+0.11-wbds(2),0.85-
wbds(2));
end
%
% See if limits are valid
%
if (sran(1)>=sran(2)),
   disp(' ')
   disp(['Lower bound ',num2str(sran(1)),...
         ' must be less than upper bound',num2str(sran(2))]);
   disp(' ')
   sran(1)=sran(2)-1;
end
%
% Clip data
%
i=min(find(xvec>=sran(1)));
j=max(find(xvec<=sran(2)));
%
% Check to see if user set window bounds manually...
%
if (sbds==[0 0 0 0]),
  sbds=[sran(1) sran(2) min(yvec(i:j)) max(yvec(i:j))];
else
  sbds(1)=min(sran(1),sbds(1));
  sbds(2)=max(sran(2),sbds(2));
end
%
% Plot delimiters...
%
```

```
sl=pbds(1)+plenx*(wbds(1)-0.13+wbds(3)*(sran(1)-sbds(1))/(sbds(2)-
sbds(1)))/0.7775;
sr=pbds(1)+plenx*(wbds(1)-0.13+wbds(3)*(sran(2)-sbds(1))/(sbds(2)-
sbds(1)))/0.7775;
y1=min(ymax,pbds(3)+pleny*((wbds(2)-0.11)/0.815-0.1));
y2=pbds(3)+pleny*((wbds(2)-0.11)/0.815-0.07);
y3=pbds(3)+pleny*((wbds(2)-0.11)/0.815-0.05);
plot([sran(1) sran(1) sl sl],[pbds(3) y1 y2 y3],'r-')
plot([sran(2) sran(2) sr sr],[pbds(3) y1 y2 y3],'r-')
hold off
%
% Switch axes handle for small plot...
%
gca1=gca;
gca2=axes('position',wbds);
plot(xvec,yvec)
axis(sbds)
axes(gca1);
if (nargout>0),
  h=[gca1 gca2];
end
%
return
```

The three functions 'plotacf.m', 'plotxcf.m' and 'plotcra.m' were used to calculate and display the results of the auto-correlation and cross correlation functions and the impulse response coefficients, respectively.

```
function [acf,lim]=plotacf(y,errlim,len,code,TitleExt)

%   Function to plot normalized autocorrelation
%
%   [acf,lim]=plotacf(y,errlim,len,code,TitleExt)
%
%   Input:       y - time series data
%          errlim - = 0 --> no error limits
%                   <>0 --> error limit: 2/sqrt(length(y))
%             len - = 0 --> number of lags: length(y)/2
%                   > 0 --> uses 'len' number of lags
%            code - = 0 --> no plot
%                   <>0 --> plot results
%         TitleExt - Optional plot title extension (string)
%
%   Output:    acf - autocorrelation values
%              lim - 95% confidence limit
%
%   Eb Mueller, March 1996

if nargin < 4
  disp('A minimum of 4 input arguments are required for this...
        function:');
```

```
  help plotacf;
  return
end

acf=xcov(y);
l=length(y);
acf=acf/acf(1);
r=1:2*(l-1);
lim=2/sqrt(l);
rl=1:length(r);
N=length(rl);
acf=acf(r);

if len>0 & len<N
  rl=N/2+1:N/2+len+1;
  acf=acf(rl);
  N=0:len;
else
  len=N/2-1;
  rl=N/2+1:N/2+1+len;
  acf=acf(rl);
  N=0:len;
end

if code>0

  if errlim>0
    plot(N,acf,...
         N,acf,'o',...
         N, lim*ones(size(N)),'r:',...
         N,-lim*ones(size(N)),'r:',...
         N, zeros(size(N)),'k-');
  else
    plot(N,acf,N,acf,'o');
  end

  if nargin == 4
    TitleExt='';
  end

  axis([0 len min(min(acf),-lim-0.1) 1.1]);
  title(['Normalized Autocorrelation Function ',TitleExt]);
  ylabel('ACF');
  xlabel('Lag');
end

return


function [xcf,lim]=plotxcf(y,u,errlim,len,code,TitleExt)

%   Function to plot normalized crosscorrelation
%
```

```
%   [xcf,lim]=plotxcf(y,u,errlim,len,code,TitleExt)
%
%   Input:        y - plant output data
%                 u - plant input data
%           errlim - = 0 --> no error limits
%                    <>0 --> error limit: 2/sqrt(length(y))
%              len - = 0 --> number of lags: length(y)/2
%                    > 0 --> uses 'len' number of lags
%             code - = 0 --> no plot
%                    <>0 --> plot results
%         TitleExt - Optional plot title extension (string)
%
%   Output:    xcf - crosscorrelation values
%              lim - 95% confidence limit
%
%   Eb Mueller, March 1996

if nargin < 5
  disp('A minimum of 5 input arguments are required for this...
       function:');
  help plotxcf;
  return
end

if length(u) ~= length(y)
   disp('The two data series must be the same size!');
   return
end

xcf=xcov(u,y);
l=length(y);
xcf=xcf/max(abs(xcf));
r=1:2*(l-1);
lim=2/sqrt(l);
rl=1:length(r);
N=length(rl);
xcf=xcf(r);

if len>0 & len<N
  rl=N/2+1:N/2+len+1;
  xcf=xcf(rl);
  N=0:len;
else
  len=N/2-1;
  rl=N/2+1:N/2+len+1;
  xcf=xcf(rl);
  N=0:len;
end

min_y=min(min(xcf),-lim-0.1);
max_y=1.1;

if code>0
```

```
  clf reset;
  if errlim>0
    plot(N,xcf,...
         N,xcf,'o',...
         N, lim*ones(size(N)),'r:',...
         N,-lim*ones(size(N)),'r:',...
         N, zeros(size(N)),'k-');
  else
    plot(N,xcf,N,xcf,'o');
  end

  if nargin == 5
    TitleExt='';
  end

  axis([0 len min_y max_y]);
  title(['Normalized Crosscorrelation Function ',TitleExt]);
  ylabel('XCF');
  xlabel('Lag');
end

return



function [ir,r,cl]=plotcra(y,u,na,len,plot,TitleExt)

%   Function to plot impulse response estimate
%
%   [ir,r,cl]=plotcra(y,u,na,len,plot,TitleExt)
%
%   Input:
%        y - plant output data
%        u - plant input data
%       na - = 0 --> no prewhitening
%            > 0 --> order of prewhitening filter (Default = 10)
%      len - = 0 --> number of lags: 20 (Default)
%            > 0 --> uses 'len' number of lags
%     plot - = 0 --> no plot
%            = 1 --> plot impulse response coefficients (Default)
%            = 2 --> plot all columns of r
%     TitleExt - Optional plot title extension (string)
%
%   Output:
%       ir - impulse response coefficients
%   r(:,1) - lag indices
%   r(:,2) - covariance function of y (prewhitened if na > 0)
%   r(:,3) - covariance function of u (prewhitened if na > 0)
%   r(:,4) - correlation function between (prewhitened if na > 0)
%            u and y (positive lags indicate influence from u to y)
%             cl - 99% confidence limit
%
%   Eb Mueller, March 1996
```

```
if nargin < 6
  TitleExt='';
end

if nargin < 5
  plot=1;
end

if nargin < 4
  len=20;
end

if nargin < 3
  na=10;
end

if nargin < 2
  disp('A minimum of 2 input arguments are required for this...
        function:');
  help plotcra;
  return
end

if isempty(plot)
  plot=1;
end

if isempty(len)
  len=20;
end

if isempty(na);
  na=10;
end

clf reset;
[ir,r,cl]=cra([y u],len+1,na,plot);

axis([0 len min(-cl-0.05,min(ir)-0.05) max(cl+0.05,max(ir)+0.05)]);
title(['Impulse Response Estimation ',TitleExt]);
ylabel('Impulse Response');
xlabel('Lag');

return
```

The spectral analysis was carried out using the function 'fdist.m'.

```
function [power,freq,period]=fdist(x,t)

%  Function for calculating and plotting the frequency distribution
%    of a data set.
%
%  Input:
```

```
%        x - data vector
%        t - title for plot enclosed in single quotes ('...')
%            (if no title is specified, no plot will be generated)
%
%  Outputs:
%    power - power vector
%     freq - frequency vector
%   period - period vector
%
%   Eb Mueller, March 1996

y = fft(x);
y(1) = [];
n = length(y);

power = abs(y(1:n/2)) .^ 2;
freq = [(1:n/2)/(n/2) * 0.5]';
period = 1 ./ freq;

if nargin == 2
   clf reset;
   plot(period,power);
   ylabel('Power');
   xlabel('Period (sample intervals per cycle)');
   title(t);
end

return;
```

The 'filtord.m' function was employed for the design and evaluation of the digital data filters.

```
function [a,b]=filtord(f,m,minord,maxord)

%   Comparison of different order filters
%
%   [a,b]=filtord(f,m,minord,maxord)
%
%   Inputs:
%          f - frequency specification vector
%              range: 0 to 1, where 1.0 corresponds to
%                     one half the sampling frequency
%              value: frequency / (0.5*sample frequency)
%          m - magnitude specification vector (magnitude
%              of desired response at the points specified in f).
%     minord - minimum filter order to be evaluated
%     maxord - maximum filter order to be evaluated
%
%   Outputs:
%      b & a - matrices of numerator and denominator filter
%              coefficients, respectively.
%              first row  --> 'minord' order filter coefficients
```

```
%              second row --> 'minord+1' order filter coefficients
%               ...             ...
%              last row   --> 'maxord' order filter coefficients
%
%   See also:  yulewalk
%
%   Eb Mueller, March 1996

a=zeros(maxord-minord+1,maxord+1);
b=zeros(maxord-minord+1,maxord+1);
h=zeros(128,maxord-minord+1);
w=zeros(128,maxord-minord+1);
n=0;

for i=minord:maxord

   n=n+1;

   [bi,ai]=yulewalk(i,f,m);
   a(n,:)=[ai zeros(1,maxord-i)];
   b(n,:)=[bi zeros(1,maxord-i)];

   [hi,wi]=freqz(bi,ai,128);
   h(:,n)=hi;
   w(:,n)=wi;

end

clf reset;
plot(f,m,w/pi,abs(h));
grid;

if (maxord-minord) == 0
   s1=int2str(maxord);
   title(['Filter Response (Filter Order: ',s1,')']);
else
   s1=int2str(minord);
   s2=int2str(maxord);
   title(['Filter Response (Filter Orders: ',s1,' to ',s2,')']);
end

xlabel('f/(0.5*fs)');
ylabel('Magnitude');

return;
```

Removal of time delays from the data set was done with the function 'shift.m'.

```
function [xb,yb]=shift(xdata,xshift,ydata,yshift)

%   Function to remove time delays from process data.
%
%   [xb,yb]=shift(xdata,xshift,ydata,yshift)
```

```
%
%   Inputs:  xdata - process input data block
%           xshift - row vector with as many elements
%                    as there are columns in xdata giving
%                    the delay for each process input
%            ydata - process output data block
%           yshift - row vector with as many elements
%                    as there are columns in ydata giving
%                    the delay for each process output
%
%   Outputs:    xb - time shifted process input data block
%               yb - time shifted process output data block
%
%   Note: Select the earliest measurement in both the input and
%         output data blocks and use it as the reference point for
%         'xshift' and 'yshift'.  All delays specified in these two
%         vectors must be positive.
%
%   Eb Mueller, March 1996

[nx,mx]=size(xdata);
xs=length(xshift);
[ny,my]=size(ydata);
ys=length(yshift);

if (mx ~= xs) | (my ~= ys)
  disp('Error in input arguments');
  help shift
  return
end

if (min(xs) < 0) | (min(ys) < 0)
  disp('All delays must be positive');
  help shift
  return
end

maxdelay=max([max(xshift) max(yshift)]);

for i=1:xs
  xb(:,i)=xdata(xshift(i)+1:nx-maxdelay+xshift(i),i);
end

for i=1:ys
  yb(:,i)=ydata(yshift(i)+1:ny-maxdelay+yshift(i),i);
end

return;
```

The filtering of the data matrices used for the PLS analysis was done with the function 'datafilt.m'.

```
function [xf]=datafilt(b,a,x)

%   Function for filtering data matrix.  The filter
%   is applied to the data one column at time with
%   initial condition matching.
%
%   [xf]=datafilt(b,a,x)
%
%   Inputs:    b = filter numerator
%              a - filetr denominator
%              x - data matrix
%
%   Output:  xf - filtered data matrix
%
%   See also: filter
%
%   E. Mueller, July 1996

[row,col]=size(x);

[yy,xx]=dstep(b,a);
n=3*length(yy);

for j=1:col
   [temp,ic]=filter(b,a,x(1,j)*ones(n,1));
   xf(:,j)=filter(b,a,x(:,j),ic);
end

return
```

The adaptive simulation algorithm was implemented in the 'sim.m' function.

```
function [yp,fbeta,lambda]=sim(x,y,r,ibeta)

%   Simulation function with moving average
%     bias correction (MISO system).
%
%   Inputs:
%        x - plant input data matrix
%        y - plant output data vector
%        r - regression coefficient vector
%    ibeta - initial value for bias correction
%
%   Output:
%       yp - output prediction
%    fbeta - final value for bias correction
%   lambda - correction term trajectory
%
%   Eb Mueller, September 1996

[nx,mx]=size(x);
[ny,my]=size(y);
```

```
if nx ~= ny
   disp('Input and output data must have the same number of
observations');
   help sim;
   return;
end

if my ~= 1
   disp('This routine applies to MISO systems only');
   help sim;
   return;
end

yp = zeros(ny,1);
e = zeros(60,1);
beta = ibeta;
k = floor(nx/60);
lambda=zeros(k,1);
t=0;

for j = 1:k
   lambda(j) = beta;
   for i = 1:60
      t = t+1;
      yp(t) = x(t,:)*r+beta;
      e(i)= y(t)-yp(t);
   end
   beta = beta+sum(e)/60;
end

for i = t+1:nx
   t=t+1;
   yp(t) = x(t,:)*r + beta;
end

fbeta = beta;

return;
```

References

[1] Barry M. Wise, 1415 Wright Avenue, Richland, WA 99352; e-mail: bm_wise@pnl.gov.